HIGH-THROUGHPUT SEQUENCING-BASED INVESTIGATION OF CHRONIC DISEASE MARKERS AND MECHANISMS

EDITED BY: Hua Li, Wen-Lian Chen, Yuriy L. Orlov and Guoshuai Cai
PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-559-1
DOI 10.3389/978-2-88976-559-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

HIGH-THROUGHPUT SEQUENCING-BASED INVESTIGATION OF CHRONIC DISEASE MARKERS AND MECHANISMS

Topic Editors:
Hua Li, Shanghai Jiao Tong University, China
Wen-Lian Chen, Shanghai University of Traditional Chinese Medicine, China
Yuriy L. Orlov, I.M.Sechenov First Moscow State Medical University, Russia
Guoshuai Cai, University of South Carolina, United States

Citation: Li, H., Chen, W.-L., Orlov, Y. L., Cai, G., eds. (2022). High-throughput Sequencing-based Investigation of Chronic Disease Markers and Mechanisms. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-559-1

Table of Contents

frontiers | Frontiers in Genetics

EDITORIAL
published: 21 June 2022
doi: 10.3389/fgene.2022.922206

Editorial: High-Throughput Sequencing-Based Investigation of Chronic Disease Markers and Mechanisms

*Yuriy L. Orlov[1,2,3,4], Wen-Lian Chen[5], Marina I. Sekacheva[1], Guoshuai Cai[6] and Hua Li[7]**

[1]I. M. Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia, [2]Novosibirsk State University, Novosibirsk, Russia, [3]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, [4]Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia, [5]Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China, [6]Environmental Health Sciences, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States, [7]Bio-ID Center, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China

Keywords: high-throughput sequencing, biomarker development, chronic disease, disease mechanism, omics study, prognosis prediction

Editorial on the Research Topic

High-throughput sequencing-based investigation of chronic disease markers and mechanisms

This "High-throughput sequencing-based investigation of chronic disease markers and mechanisms" issue focuses on the genomics studies on cancer and chronic diseases. With the recent development of sequencing technology and the rapid reduction of sequencing costs, high-throughput sequencing (including second and third-generation sequencing) is revolutionizing basic life science research and clinical research. High-throughput sequencing often produces millions of sequencing reads at a time, and the alignment or assembly of these reads allows determination of various mutations at the genomic level, accurate gene expression quantification at the transcriptomic level, and identification of histone or DNA modification at the epigenomic level. The resulting accumulation of enormous multi-omics information has opened up a new era of finding effective disease markers and studying their roles in disease occurrence and development (Anashkina et al., 2021).

Using high-throughput sequencing, various markers of chronic diseases have been developed at all omics levels, which have been used for diagnosis and classification of diseases, prediction of treatment effects, and prevention of diseases (Voronova et al., 2020; Glukhov et al., 2021). The chronic diseases include cancer, heart disease, diabetes, arthritis. The quickly and massively acquired multi-omics data, together with newly developed algorithms, provide excellent opportunities for the identification of more reliable biomarkers. This Research Topic aimed at 1) developing new chronic disease markers at four levels (i.e., genome, epigenome, transcriptome, and translatome) with the help of high-throughput sequencing, and 2) delineating potential marker-related mechanisms for chronic disease occurrence or development. More specifically, this Research Topic contains contributions including:

1) Identification of novel biomarkers and prediction signatures for chronic disease detection or prognosis prediction using high-throughput sequencing;
2) Analysis the possible pathological causes of markers as well as the potential roles they play in disease initiation and development;
3) Applications of new high-throughput sequencing techniques facilitating the development of more effective biomarkers of chronic disease;

Edited and reviewed by:
Jared C. Roach, Institute for Systems Biology (ISB), United States

****Correspondence:***
Hua Li
kaikaixinxin@sjtu.edu.cn

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *17 April 2022*
Accepted: *12 May 2022*
Published: *21 June 2022*

Citation:
Orlov YL, Chen W-L, Sekacheva MI, Cai G and Li H (2022) Editorial: High-Throughput Sequencing-Based Investigation of Chronic Disease Markers and Mechanisms. Front. Genet. 13:922206. doi: 10.3389/fgene.2022.922206

4) New algorithms or tools for *in silico* identification of effective chronic disease markers based on high-throughput sequencing data.

Thus, we have organized this Research Topic to collect the papers focused on the frontiers of chronic disease markers. This Topic complements recent Research Topics "Bioinformatics of Genome Regulation" (Orlov et al., 2021a) and "Association between Individuals' Genomic Ancestry and Variation in Disease Susceptibility" in *Frontiers in Genetics* (Das et al., 2022). The later journal issue collected papers focused on genetic background and ancestry rather than on molecular mechanisms of the human diseases (Das et al., 2022).

The papers published here extend the studies presented in the *Frontiers in Genetics* Topic "Bioinformatics of Genome Regulation (https://www.frontiersin.org/research-topics/17947/bioinformatics-of-genome-regulation-volume-ii), Volume II" and the earlier "Bioinformatics of Genome Regulation and Structure \ Systems Biology (BGRS\SB)" conference series (https://www.frontiersin.org/research-topics/8383/bioinformatics-of-genome-regulation-and-systems-biology; Orlov and Baranova, 2020; Orlov et al., 2016).

In this Research Topic a total of 14 papers could be arranged by two main areas—the cancer studies, and the works on the other chronic diseases such as allergy. Lung cancer is one of the leading causes of cancer-associated death in the world. We open the Research Topic by group of papers on lung cancer.

Non-small cell lung cancer (NSCLC) comprises about 85% of all lung cancers. Chang et al. used whole-exome sequencing to explore platinum-drug resistance mutations in advanced non-small cell lung cancer. Platinum-based chemotherapy is a fundamental treatment for non-small cell lung cancer (NSCLC) patients who are not suitable for targeted drug therapies. However, most patients progressed after a period of treatment. The authors enrolled nine NSCLC patients with platinum-based chemotherapy resistance, collected circulating tumor cells from them and performed whole-exome sequencing. Lu et al. studied response of TP53-negative NSCLC Patients to atezolizumab, an immune checkpoint inhibitor. This study provides a predictor, ubiquitin-like conjugation biological process gene mutation status, for identifying NSCLC patients who may have response to atezolizumab therapy.

Lung adenocarcinoma is the most common subtype of lung cancer with heterogeneous outcomes and diverse therapeutic responses. Wu et al. have developed novel risk-score model with eight miRNA signatures for overall survival of patients with lung adenocarcinoma. The authors selected eight microRNA (miRNA) signatures in The Cancer Genome Atlas. This model can also provide new insights into the current clinical staging system and can be regarded as an alternative system for patient stratification.

Wang et al. revealed the migration-associated genes involved in antitumor effects of herbal medicine Feiyanning on lung cancer cells using RNA-Seq and ATAC-Seq analysis. Traditional Chinese medicine formula Feiyanning has been clinically administered in China for more than a decade and raised attention due to its anticancer effect in lung cancer. Using cellular and molecular assays to examine the antitumor activities in lung cancer cells this study suggested that Feiyanning formula exerted the antitumor effects by modulating the expression and chromatin accessibility levels of migration-associated genes.

Breast cancer, the most commonly diagnosed cancer in women, has posed a major threat to women's health globally. Hu et al. have compared pegylated liposomal doxorubicin (PLD) with epirubicin as adjuvant therapy for stage I-III breast cancer. Based on the large sample size and the long follow-up time of this study, the authors conclude that PLD has a similar anti-breast cancer efficacy as epirubicin while inducing lower level of cardiac toxicity in Han Chinese.

Gastric cancer is the third leading cause of cancer mortality across the world. Shi X. et al. studied CPSF6 protein in human gastric cancer. Alternative polyadenylation (APA) affects various biological functions and is involved in cancer. The authors found that 19 of 20 core APA factors were upregulated in gastric cancer tissues.

Esophageal cancer is the eighth most common cancer and the sixth leading cause of cancer death worldwide. Pang et al. estimated prognostic value of immune-related multi-IncRNA signatures associated with tumor microenvironment in esophageal cancer. The authors analyzed the tumor microenvironment for two subtypes of esophageal cancer, identified two multi-lncRNA signatures predictive for the prognosis, and explored the possibility of the signatures to forecast drug susceptibility.

The liver cancer ranks sixth in the number of new cases of malignant tumors worldwide and is the third leading cause of cancer death in the world. Hepatocellular carcinoma is the most common pathological type of primary liver cancer, accounting for about 90%. Shi H. et al. constructed a prognostic model based on the peptidyl prolyl cis–trans isomerase gene signature and explored it in patients with hepatocellular carcinoma.

Knyazev et al. studied hypoxia-related markers in inflammatory bowel disease. The authors detected expression activation of ITGA5 and PLAUR genes encoding integrin α5 and urokinase-type plasminogen activator receptor in inflammatory bowel disease specimens. The interaction of these molecules can activate cell migration and regenerative processes in the epithelium. These genes can serve as markers of inflammatory bowel disease progression and intestinal hypoxia.

The extracellular matrix (ECM) and cellular receptors constitute one of the crucial pathways involved in colorectal cancer progression and metastasis. Nersisyan et al. studied ECM–receptor regulatory network in colorectal cancer. The authors evaluated the prognostic information concentrated in the genes from ECM–receptor network using transcription factor and miRNA data and constructed two prognostic signatures.

Abnormal expression and regulation of non-coding RNA are involved in a variety of human diseases. Series of papers presenting miRNA and ncRNA studies focused on different groups of chronic diseases. Senile osteoporosis has recently become a major chronic metabolic bone disease in the world. Geng et al. aims to identify differentially expressed mRNAs and ncRNAs in senile osteoporosis patient-derived bone mesenchymal stem cells. By constructing the

circRNA–miRNA–mRNA regulatory network, they found a ceRNA network (circRNA008876-miR-150-5p-mRNA) that could play an important role in senile osteoporosis.

The prevalence of allergic diseases has been increasing worldwide over the past 60 years, affecting about 30% of the global population. Qiao et al. described six male patients from unrelated families with a triad symptom of progressive postnatal slow growth, allergies, and fatty liver. The authors show association of imprinted gene SLC22A18 (solute carrier family 22 member 18) with this syndrome of variable allergy, short stature, and fatty liver. New variant in the promoter region of SLC22A18 is potentially associated with this syndrome.

Fei et al. studied potential associations and detect the underlying impact of single-nucleotide polymorphisms (SNPs) on proteinuria in kidney transplant recipients. The study suggested that the mutation of rs3804099 on the TLR2 gene was significantly related to the generation of proteinuria after kidney transplantation.

Wang et al. analyzed the expression and prognostic value of gamma-glutamyl-transferase 5 (GGT5) and its correlation with immune cell infiltration in gastric cancer. GGT5 may serve as a promising prognostic biomarker and a potential immunological therapeutic target for gastric cancer, since it is associated with immune cell infiltration in the tumor microenvironment.

Overall, we are proud of the Research Topic at *Frontiers in Genetics* we collated. We hope that the readers will find this collection stimulating and consider participation in upcoming conferences and journal issues in this area (https://bgrssb.icgbio.ru/2022/). Note also the thematic issue "Medical Genetics, Genomics and Bioinformatics" on the sequencing technologies applications in medical genetics (https://www.mdpi.com/journal/ijms/special_issues/Medical_Genetics_2022) and the continuing topic on gene expression mechanisms (https://www.frontiersin.org/researchtopics/17947/bioinformatics-of-genome-regulation-volume-ii) (Orlov et al., 2021a; 2021b). The complementary special issue at *Life* journal (https://www.mdpi.com/journal/life/special_issues/identification_HTS) continues collection of papers on the diseases markers and underlying molecular mechanisms (Snezhkina et al., 2021).

AUTHOR CONTRIBUTIONS

HL, W-LC, GC, and YO organized the Research Topic as guest editors, supervised the reviewing of the manuscripts, MS critically contributed both to the extension of the Topic and the reviewing process. All the authors wrote this Editorial paper. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The guest editors are grateful to all the authors contributing to this special issue papers collection and thank all the reviewers who helped improve the manuscripts. The publication has been prepared with the support of the RUDN University Scientific Projects Grant System, project R.3-2022-ins (YO).

REFERENCES

Anashkina, A. A., Leberfarb, E. Y., and Orlov, Y. L. (2021). Recent Trends in Cancer Genomics and Bioinformatics Tools Development. *Int. J. Mol. Sci.* 22, 12146. doi:10.3390/ijms222212146

Das, R., Tatarinova, T. V., Galieva, E. R., and Orlov, Y. L. (2022). Editorial: Association between Individuals' Genomic Ancestry and Variation in Disease Susceptibility. *Front. Genet.* 13, 831320. doi:10.3389/fgene.2022.831320

Glukhov, A., Potoldykova, N., Taratkin, M., Gordeev, S., Polyakovsky, K., Laukhtina, E., et al. (2021). Detection of Urothelial Bladder Cancer Based on Urine and Tissue Telomerase Activity Measured by Novel RT-TRAP-2pcr Method. *J. Clin. Med.* 10 (5), 1055. doi:10.3390/jcm10051055

Orlov, Y. L., Anashkina, A. A., Klimontov, V. V., and Baranova, A. V. (2021b). Medical Genetics, Genomics and Bioinformatics Aid in Understanding Molecular Mechanisms of Human Diseases. *Int. J. Mol. Sci.* 22 (18), 9962. doi:10.3390/ijms22189962

Orlov, Y. L., Anashkina, A. A., Tatarinova, T. V., and Baranova, A. V. (2021a). Editorial: Bioinformatics of Genome Regulation, Volume II. *Front. Genet.* 12, 795257. doi:10.3389/fgene.2021.795257

Orlov, Y. L., and Baranova, A. V. (2020). Editorial: Bioinformatics of Genome Regulation and Systems Biology. *Front. Genet.* 11, 625. doi:10.3389/fgene.2020.00625

Orlov, Y. L., Baranova, A. V., Hofestädt, R., and Kolchanov, N. A. (2016). Computational Genomics at BGRS\SB-2016: Introductory Note. *BMC genomics* 17 (Suppl. 14), 996. doi:10.1186/s12864-016-3350-6

Snezhkina, A., Pavlov, V., Dmitriev, A., Melnikova, N., and Kudryavtseva, A. (2021). Potential Biomarkers of Metastasizing Paragangliomas and Pheochromocytomas. *Life (Basel, Switz.* 11 (11), 1179. doi:10.3390/life11111179

Voronova, V., Glybochko, P., Svistunov, A., Fomin, V., Kopylov, P., Tzarkov, P., et al. (2020). Diagnostic Value of Combinatorial Markers in Colorectal Carcinoma. *Front. Oncol.* 10, 832. doi:10.3389/fonc.2020.00832

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 07 September 2021
doi: 10.3389/fgene.2021.723670

ctDNA-Profiling-Based UBL Biological Process Mutation Status as a Predictor of Atezolizumab Response Among *TP53*-Negative NSCLC Patients

Jun Lu[1,2,3†], Yanwei Zhang[1†], Yuqing Lou[1†], Bo Yan[1,4], Benkun Zou[1], Minjuan Hu[1], Yanan Wang[1], Ya Chen[1], Zhengyu Yang[1], Huimin Wang[1*], Wei Zhang[1*] and Baohui Han[1,2,3*]*

[1] Department of Pulmonary Medicine, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [2] Shanghai Institute of Thoracic Oncology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [3] Translational Medical Research Platform for Thoracic Oncology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [4] Clinical Research Center, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China

OPEN ACCESS

Edited by:
Guoshuai Cai,
University of South Carolina,
United States

Reviewed by:
Jun Wu,
East China Normal University, China
Zhuo Wang,
Fudan University, China

***Correspondence:**
Jun Lu
lujun512@yahoo.com
Huimin Wang
chestwhm@126.com
Wei Zhang
zhwei2002@hotmail.com
Baohui Han
18930858216@163.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 11 June 2021
Accepted: 09 August 2021
Published: 07 September 2021

Citation:
Lu J, Zhang Y, Lou Y, Yan B, Zou B, Hu M, Wang Y, Chen Y, Yang Z, Wang H, Zhang W and Han B (2021) ctDNA-Profiling-Based UBL Biological Process Mutation Status as a Predictor of Atezolizumab Response Among TP53-Negative NSCLC Patients. Front. Genet. 12:723670. doi: 10.3389/fgene.2021.723670

Atezolizumab, an immune checkpoint inhibitor, has been approved for use in clinical practice in non-small cell lung cancer (NSCLC) patients, but potential biomarkers for response stratification still need further screening. In the present study, a total of 399 patients with high-quality ctDNA profiling results were included. The mutation status of ubiquitin-like conjugation (UBL) biological process genes (including *ABL1*, *APC*, *LRP6*, *FUBP1*, *KEAP1*, and *TOP2A*) and clinical information were further integrated. The results suggested that the patients with the clinical characteristics of male or history of smoking had a higher frequency of UBL mutation positivity [UBL (+)]; the patients who were UBL (+) had shorter progression-free survival (PFS) (1.69 vs. 3.22 months, $p = 0.0007$) and overall survival (8.61 vs. 16.10 months, $p < 0.0001$) than those patients with UBL mutation negativity [UBL (–)]; and more promising predictive values were shown in the smoker subgroup and ≤ 3 metastasis subgroup. More interestingly, we found the predictor has more performance in *TP53*-negative cohorts [training in an independent POPLAR and OAK cohorts ($n = 200$), and validation in an independent MSKCC cohort ($n = 127$)]. Overall, this study provides a predictor, UBL biological process gene mutation status, not only for identifying NSCLC patients who may respond to atezolizumab therapy but also for screening out the potential NSCLC responders who received other immune checkpoint inhibitors.

Keywords: UBL, biomarker, immune checkpoint inhibitors, atezolizumab, NSCLC

INTRODUCTION

Non-small cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancers (Lu et al., 2019b; Zhang et al., 2020; Lou et al., 2021). Standard therapeutic regimens for first-line therapy have been recommended for NSCLC patients according to the guidelines of the National Comprehensive Cancer Network (Ettinger et al., 2021). Patients harboring EGFR mutations are recommended to receive tyrosine kinase inhibitors (TKIs) such as osimertinib and gefitinib

(Zhao et al., 2019; Ettinger et al., 2021); patients harboring ALK mutations are recommended to receive TKIs such as alectinib and crizotinib (Peters et al., 2017; Shaw et al., 2020). However, patients without driver gene mutations usually receive chemotherapy or immunotherapy (Gadgeel et al., 2020; Ettinger et al., 2021). Regarding second-line therapy for NSCLC patients, immunotherapy is potentially suitable for patients with PD-L1 expression (Fehrenbacher et al., 2016; Rittmeyer et al., 2017).

The POPLAR study and OAK studies demonstrated that NSCLC patients who received atezolizumab (one of immune checkpoint inhibitors) as second-line therapy had significantly prolonged overall survival (OS) compared with docetaxel patients, regardless of PD-L1 expression or histology (Fehrenbacher et al., 2016; Rittmeyer et al., 2017). After that, Gandara et al. (2018) found that ctDNA profiling was a potential technology to be used for atezolizumab response stratification in the above NSCLC patients. Furthermore, our previous study suggested that the ctDNA profiling potentially provides more information for immunotherapeutic stratification (Nie et al., 2020). Although an increasing number of studies have introduced that genetic profiling can be used for predicting the efficacy of immunotherapy (Chan et al., 2019; Fabrizio et al., 2021; McGrail et al., 2021), further screening of gene cluster-based biomarkers for immunotherapy is still an urgent problem that needs to be resolved.

Ubiquitin-like conjugation (UBL) biological processes play an important role in cancer development, progression, and therapy. However, the underlying role in immunotherapy is still unclear. In the present study, we found that UBL-enriched gene (*ABL1*, *APC*, *LRP6*, *FUBP1*, *KEAP1*, and *TOP2A*) mutation significantly affected atezolizumab efficacy as second-line therapy in NSCLC patients, and we sought to identify a biomarker for potential use in clinical practice in the future.

MATERIALS AND METHODS

Patients

In the present study, all blood samples were collected from enrolled NSCLC patients from the POPLAR study (NCT01903993) and OAK study (NCT02008227). All clinical trials were performed according to Good Clinical Practice guidelines and those of the Declaration of Helsinki. The purpose of blood sample collection was completely explained to patients or their family members, and the signed informed consent was obtained. In the POPLAR study, 144 of 287 patients received atezolizumab therapy, and 425 of 850 patients in the OAK

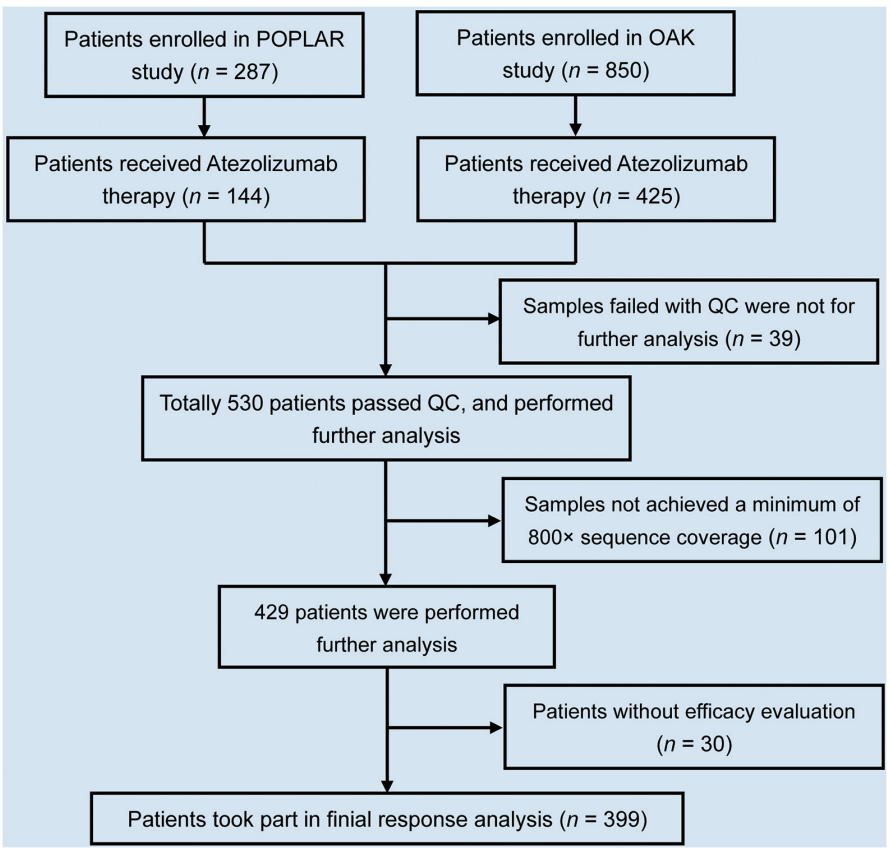


FIGURE 1 | Flowchart of NSCLC patient selection in this study. In total, 1,137 NSCLC patients enrolled in POPLAR and OAK studies. Of 1,137 patients, 569 received atezolizumab therapy. Of 569, 399 patients took part in final analysis after selection.

study received atezolizumab. Collectively, 569 patients were preliminarily selected in the present study.

ctDNA Profiling

The collected baseline plasma samples from all 569 atezolizumab-treated NSCLC patients underwent uniform procedures for ctDNA mutation calling. The methods for sample collection, storage conditions, cell-free DNA (cfDNA) extraction, library construction, sequencing, analysis, and mutation calling were performed according to a previously published article (Gandara et al., 2018). The clinical information for each patient was downloaded from the online database.[1]

Screening

In total, 569 NSCLC patients had complete clinical information and baseline blood samples. During library construction, 39 of 569 samples failed quality control (including the samples with cfDNA < 20 ng, and the samples failed with templated extension). After sequencing, 101 of 530 ctDNA profiles that did not achieve a minimum of 800× sequence coverage were excluded. Thirty of 429 patients without a definite efficacy evaluation were excluded. Therefore, in this study, 399 patients were selected for final analysis (**Figure 1**).

UBL Biological Process

We found 394 cancer-related genes with mutations in all ctDNA samples from 399 patients. Then, the list of these genes was uploaded to the DAVID database for biological process enrichment analysis. UBL biological processes were significantly enriched and mainly included six genes (*ABL1*, *APC*, *LRP6*, *FUBP1*, *KEAP1*, and *TOP2A*).

Kaplan–Meier Curve Analysis

This method performed as our previous studies (Lu et al., 2019a,c; Chu et al., 2021). The 399 patients who harbored any mutation in UBL biological process genes (*ABL1*, *APC*, *LRP6*, *FUBP1*, *KEAP1*, and *TOP2A*) from ctDNA profiling were defined as UBL (+). A patient who harbored two or more gene mutations of the abovementioned six genes was defined as "UBL (+) harboring 2 or more gene mutations." If ctDNA profiling detected no mutation in the above UBL biological process genes, the patient was defined as UBL (–). Based on the UBL prediction, the next procedure was stratifying progression-free survival (PFS) and OS by using GraphPad Prism 5 software.

Subgroup Analysis

According to the different clinical characteristics and driver gene mutation status, we classified the 399 patients into male and female, non-smoker and smoker, non-lung squamous carcinoma (non-LUSC) and LUSC, Asian and White, EGFR (+) and KRAS (+), Eastern Cooperative Oncology Group (ECOG) score = 0 and ECOG score = 1, and metastases ≤ 3 and metastases > 3. Kaplan–Meier curve analysis was performed to calculate the median PFS and median OS and the corresponding log-rank *p*-value. Hazard ratio (HR) was calculated by use of Cox proportional hazards model.

Validation Analysis

According to the *TP53* mutation status, we classified the 399 patients into *TP53* mutation-positive NSCLC patients and *TP53* mutation-negative NSCLC patients. Similar UBL-based stratification analysis was performed on the NSCLC patients with *TP53* mutation and the NSCLC patients without *TP53* mutation, respectively. For the validation cohort, we selected the *TP53* mutation-negative NSCLC patients who received at least one dose of immunotherapy (atezolizumab, avelumab, durvalumab, ipilimumab, nivolumab, pembrolizumab, or tremelimumab) from the Memorial Sloan Kettering Cancer Center (MSKCC). Different from the POPLAR and OAK cohorts, the mutation information of the NSCLC patients from the MSKCC cohort was derived from tumor tissue DNA (ttDNA). Furthermore, the patients undergo genomic profiling with the Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) panel. The patients from the MSKCC cohort just provided the OS information. Therefore, the predictive value of UBL for OS stratification was performed in the validation cohort.

Statistical Analysis

The Mann–Whitney *U*-test was performed to compare the age difference between the UBL (+) cohort and UBL (–) cohort.

TABLE 1 | Demographic data of 399 advanced NSCLC patients who received atezolizumab therapy.

Characteristic	UBL (+)	UBL (-)	*p*-value
Age (median, range)	62.5 (41–78)	62.7 (39–82)	0.7894[a]
Gender (%)			0.0148[b]
Male	68 (74%)	184 (60%)	
Female	24 (26%)	123 (40%)	
Smoking (%)			0.0030[b]
Non-smoker	7 (8%)	65 (21%)	
Smoker	85 (92%)	242 (79%)	
Pathology (%)			0.5877[b]
Non-LUSC	68 (74%)	218 (71%)	
LUSC	24 (26%)	89 (29%)	
Race (%)			0.1118[b]
Asian	14 (15%)	70 (23%)	
White	70 (76%)	210 (68%)	
Others	8 (9%)	27 (9%)	
Driver gene status (%)			
EGFR (+)	4 (4%)	49 (16%)	0.0040[b]
KRAS (+)	7 (8%)	20 (7%)	0.7140[b]
EML4 (+)	0	2 (>1%)	0.4377[b]
Metastases (%)			0.1299[b]
≤3	58 (63%)	219 (71%)	
>3	34 (37%)	88 (29%)	
ECOG (%)			0.9443[b]
0	32 (35%)	108 (35%)	
1	60 (65%)	199 (65%)	

[a]*Mann–Whitney U-test.*
[b]*Chi-square test.*

[1]https://clinicalstudydatarequest.com/Study-Sponsors/Study-Sponsors-Roche.aspx

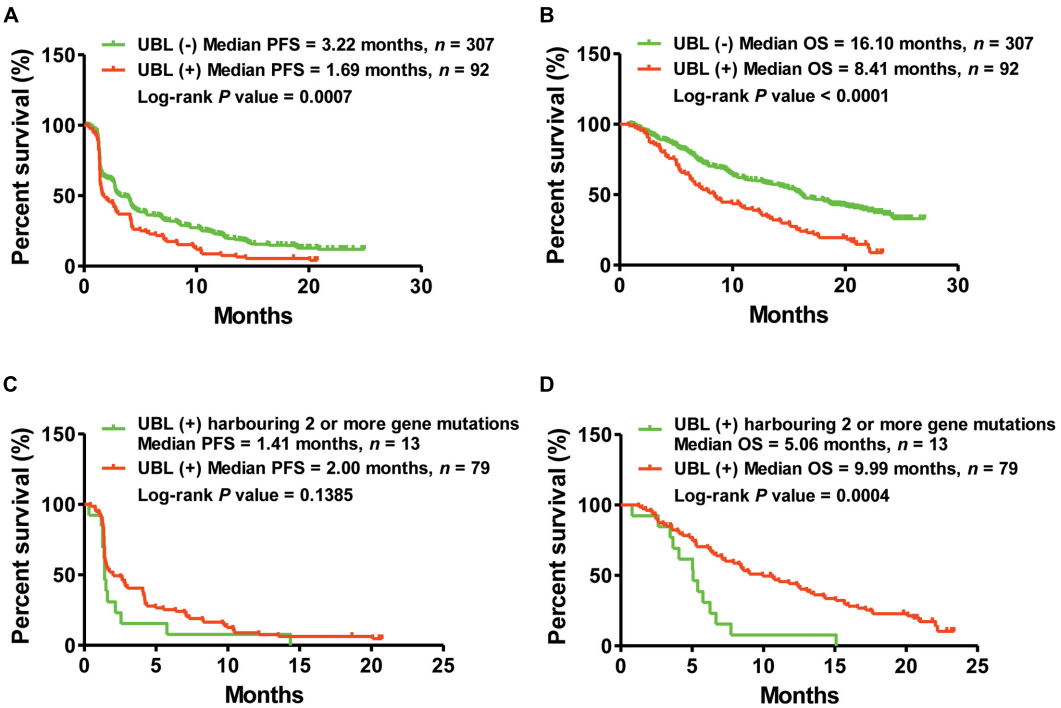


FIGURE 2 | Kaplan–Meier curve analysis of PFS and OS via the predictor of UBL. **(A)** NSCLC patients harboring UBL (+) received median PFS of 1.69 months, while those patients harboring UBL (–) received median PFS of 3.22 months. **(B)** NSCLC patients harboring UBL (+) received median OS of 8.41 months, while those patients harboring UBL (–) received median OS of 16.10 months. **(C)** UBL (+) NSCLC patients harboring two or more gene mutations received median PFS of 1.41 months, while those UBL (+) patients harboring single gene mutation received median PFS of 2.00 months. **(D)** UBL (+) NSCLC patients harboring two or more gene mutations received median OS of 5.06 months, while those UBL (+) patients harboring single gene mutation received median OS of 9.99 months.

The chi-square test was performed to compare the differences of other clinical characters. The log-rank test was used to compare Kaplan–Meier curves during the stratification analysis. The HRs and exact 95% confidence intervals (CIs) are reported. Differences were considered significant at $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

RESULTS

In this study, we identified NSCLC patients harboring UBL biological process gene (*ABL1*, *APC*, *LRP6*, *FUBP1*, *KEAP1*, and *TOP2*) mutations, and this information could be used to stratify patients who might respond to the immune checkpoint inhibitor atezolizumab. Based on our rigorous screening procedure, 399 patients were finally included for screening the responsive biomarker (**Figure 1**). The analysis of clinical characteristics indicated that there was no association between UBL gene mutation status and age, pathological status, race, driver gene (*EGFR*, *ALK*, and *EML4*) status, metastasis site number, and ECOG score. However, there was a significant increase in UBL gene mutation frequency in the male patients and the patients with smoking history (**Table 1**). These results suggested that there is a bias in the mutation of UBL biological process genes in NSCLC patients. Whether this phenomenon can be used as a predictor of immunotherapy response is still unclear.

Here, we classified 399 atezolizumab-treated NSCLC patients into two cohorts: those who were UBL (+) and those who were UBL (–). Kaplan–Meier curve analysis suggested that the

TABLE 2 | Subgroup response analysis using the biomarkers of UBL in atezoluzimab-treated patients from OAK and POPLAR cohorts.

	Median PFS (months)			Median OS (months)		
	UBL (+)	UBL (–)	*p*-value[a]	UBL (+)	UBL (–)	*p*-value[a]
Male	1.59	2.86	0.0076	7.89	15.47	<0.0001
Female	2.73	4.01	0.0904	10.04	17.15	0.0314
Non-smoker	1.38	2.79	0.0977	9.99	17.97	0.3255
Smoker	1.74	4.01	0.0009	8.28	15.77	<0.0001
Non-LUSC	1.53	3.19	0.0027	8.28	18.04	<0.0001
LUSC	2.27	3.78	0.0952	8.43	10.05	0.0283
Asian	2.00	2.86	0.7172	20.90	21.26	0.4350
White	1.61	3.61	0.0003	7.29	16.00	<0.0001
EGFR (+)	2.17	2.76	0.6583	16.54	14.23	0.5805
KRAS (+)	1.41	4.07	0.1848	9.00	18.48	0.1633
ECOG = 0	2.78	4.01	0.0691	12.90	22.47	0.0022
ECOG = 1	1.54	2.89	0.0051	6.67	14.88	<0.0001
>3 Metastases	2.27	2.48	0.8028	10.74	10.94	0.2112
≤3 Metastases	1.58	4.17	<0.0001	7.72	18.60	<0.0001

[a]*Log-rank (Mantel–Cox) test.*

NSCLC patients who were UBL (+) had shorter PFS (UBL (+) vs. UBL (–) = 1.69 vs. 3.22 months, log-rank p-value = 0.0007) than the UBL (–) cohort (**Figure 2A**). Regarding OS analysis, the results showed a more significant difference: the patients in the UBL (–) cohort received more OS benefit from atezolizumab therapy (UBL (–) vs. UBL (+) = 16.10 vs. 8.41 months, log-rank p-value < 0.0001) than the patients in the UBL (+) cohort (**Figure 2B**). More interestingly, in the UBL (+) cohort, the patients who defined as UBL (+) harboring two or more gene mutations received shorter PFS and OS from atezolizumab therapy (PFS: 1.41 vs. 2.00 months, log-rank p-value = 0.1385; OS: 5.06 vs. 9.99 months, log-rank p-value = 0.0004) (**Figures 2C,D**). These results indicated that the mutation status of the UBL biological process genes could potentially be used as a predictor of response to atezolizumab as second-line therapy in NSCLC patients.

To further understand the performance of UBL status in subgroups, we first analyzed the sex-induced response difference for NSCLC patients who received atezolizumab. Male NSCLC patients who were UBL (+) had significantly shorter PFS and OS than male patients who were UBL (–), while there was no significant difference between female patients who were UBL (+) and female patients who were UBL (–). UBL (+) patients with a history of smoking had significantly shorter PFS and OS than UBL (–) patients with a history of smoking. For those patients without a smoking history, there was no significant difference between UBL (+) patients and UBL (–) patients. Regarding the non-LUSC subgroup, patients who were UBL (+) had significantly shorter PFS and OS than male patients who were UBL (–) (**Table 2**).

For LUSC patients, UBL (+) patients had a shorter OS than UBL (–) patients, while there was no difference when comparing the PFS between these two cohorts. Neither UBL (+) nor UBL (–) labeled in Asian patients, and there was no significant PFS or OS outcome difference when patients received atezolizumab therapy. Patients harboring EGFR mutations or KRAS mutations combined with UBL biological process gene mutations who received atezolizumab therapy had PFS and OS outcomes that were similar to those of patients without UBL biological process gene mutations. Regarding the ECOG score = 1 subgroup, patients who were UBL (+) had significantly shorter PFS and OS than those without UBL (–). For the ECOG score = 0 subgroup, UBL (+) patients had a shorter OS than UBL (–) patients, while there was no difference when comparing the PFS between the two cohorts. Interestingly, UBL significantly stratified responders from non-responders in the ≤ 3 metastasis subgroup (**Table 2**). These results suggested that UBL biological process gene mutation status could potentially be used as a predictor of response to atezolizumab therapy in NSCLC patients, especially in the male, smoker, non-LUSC, White, ECOG score = 1, and ≤ 3 metastasis subgroups. In addition, our analysis focused on HR indicated that the UBL status predictor remarkably distinguished patients with an atezolizumab response and reduced risk of death in the White subgroup and the ≤ 3 metastasis subgroup (**Figure 3**).

The frequency of *TP53* mutation accounts for the highest grade (about 50%) in NSCLC. Whether the predictor of UBL plays a differential role between *TP53*-positive NSCLC and *TP53*-negative NSCLC is still unclear. To further understand the performance of UBL status in *TP53* mutation-based subgroups, here we analyzed the *TP53* mutation-induced response difference for NSCLC patients who received atezolizumab. As shown in **Figure 4A**, there is no significant distinguishing values of the predictor UBL among the NSCLC patients

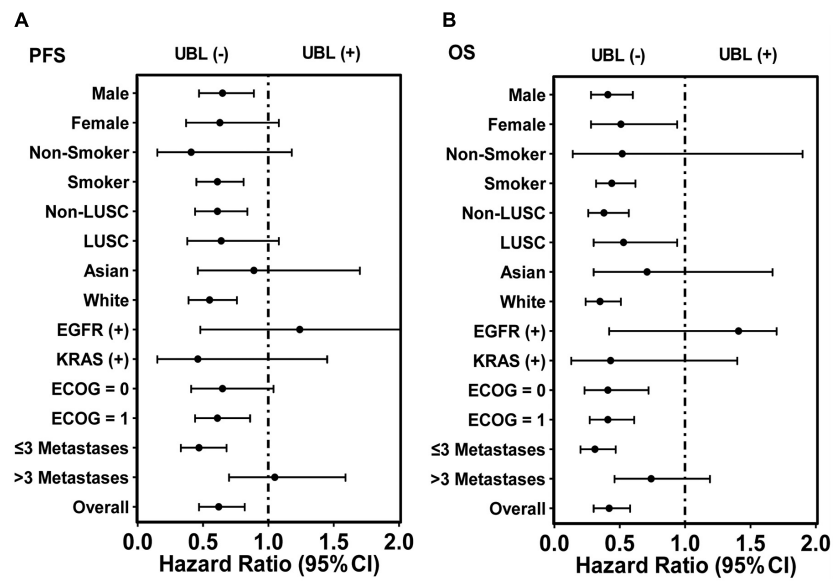


FIGURE 3 | HR analysis of overall 399 patients and corresponding subgroups *via* the predictor UBL-based stratification. **(A)** The predictor UBL potentially decreased HR in nearly all subgroups except the subgroups of EGFR (+) and > 3 metastases when PFS analysis was performed. **(B)** The predictor UBL potentially decreased HR in all subgroups except the subgroup of EGFR (+) when OS analysis was performed.

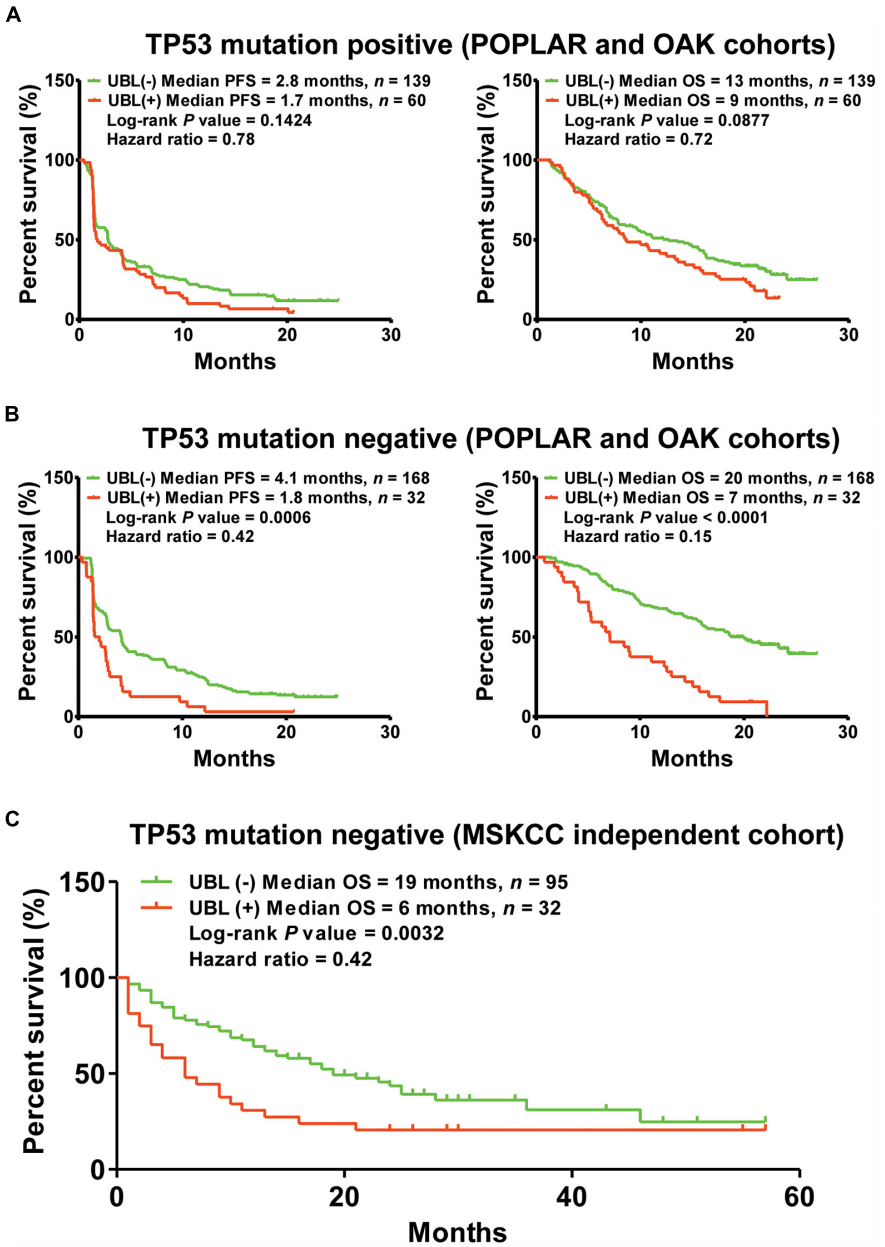


FIGURE 4 | The predictor of UBL for immunotherapeutic stratification in *TP53* mutation-based subgroups. **(A)** The predictive values of UBL in the NSCLC patients harboring *TP53* mutation. **(B)** The predictive values of UBL in the NSCLC patients without harboring *TP53* mutation. **(C)** An independent MSKCC cohort was used to validate the predictive value of UBL in the *TP53*-negative NSCLC patients who received immunotherapy.

harboring *TP53* mutation. Interestingly, the predictor UBL showed more predictive values in the *TP53* negative NSCLC patients than that in overall NSCLC patients (**Figures 3**, **4B**). To validate whether the predictor UBL can be validated in an independent cohort, we used the NSCLC patients who received immunotherapy from MSKCC center as the validation cohort. Results demonstrated that the *TP53* mutation-negative NSCLC patients with benefited OS outcome can be screened out significantly *via* the predictor UBL-based stratification (**Figure 4C**).

DISCUSSION

Immunotherapy has changed clinical practice in NSCLC (Eguren-Santamaria et al., 2020; Pinheiro et al., 2020; Ettinger et al., 2021). With the clinical application of immune checkpoint inhibitors in NSCLC, an increasing number of clinical problems have surfaced (de Miguel and Calvo, 2020; Haanen et al., 2020; Ramos-Casals et al., 2020; Shankar et al., 2020). Among these emerging clinical problems, how to screen out responders using a predictor is always confusing for clinical physicians and scientists

(Darvin et al., 2018; de Miguel and Calvo, 2020). Therefore, in the present study, we sought to identify a ctDNA profiling-based predictor that might be used to stratify responders to the immune checkpoint inhibitor atezolizumab as second-line therapy among NSCLC patients.

Previous studies have demonstrated that genetic profiling can potentially be used as a biomarker for response to immune checkpoint inhibitors (Gandara et al., 2018; Samstein et al., 2019; Alborelli et al., 2020). However, an increasing number of scientists have provided different viewpoints about the usage of genetic profiling for predicting therapeutic efficacy in immunotherapy (Fabrizio et al., 2021; McGrail et al., 2021). With the development of technology, blood-based next-generation sequencing has opened a new field of view for biomarkers predicting response to immune checkpoint inhibitors (Gandara et al., 2018). Our previous study showed that the ctDNA profiling potentially provides more information for immunotherapeutic stratification (Nie et al., 2020). Samstein et al. (2019) and McGrail et al. (2021) demonstrated that genetic profiling plays different roles in different cancer types. Furthermore, the biomarkers including PD-L1 and microsatellite instability also have some questions that need to be resolved (Chang et al., 2018; Gandara et al., 2018). Therefore, predictor discovery for immunotherapy has just started, and there is much unknown information that needs to be explored.

In this study, we provided evidence that UBL could be used as a predictor to screen out responders from non-responders among NSCLC patients who received atezolizumab as second-line therapy. Our results indicated that NSCLC patients who were male and had a smoking history had a higher frequency of being UBL (+). This is a very interesting phenomenon. Although there is not enough evidence to confirm the association of the above characteristics, we still have a reason to speculate that smoking potentially contributes to UBL biological process gene mutations according to previously reported relationships between smoking and genetic variation (Nagahashi et al., 2018). Further analysis demonstrated that NSCLC patients who were UBL (+) had shorter PFS and OS than patients who were UBL (–). Either PFS analysis or OS analysis showed very promising results for UBL biological process gene mutation status to be able to significantly screen out responders from non-responders.

The roles of UBL biological process gene mutation status in screening out responders in subgroups contributed an important composition in the present study. According to our results, 70.4% of patients were White, and 69.4% of patients had ≤ 3 metastases among all 399 NSCLC patients. Some of the bright points in the subgroup analysis are that UBL biological process gene mutation status can significantly distinguish responders and non-responders when used in the above subgroups. However, why this phenomenon occurred still requires further study.

Based on existing evidence, there may be great differences in tumor biology between patients with NSCLC harboring *TP53* mutations and those without *TP53* mutations (Mogi and Kuwano, 2011; Jamal-Hanjani et al., 2017; Bailey et al., 2018; Birkbak and McGranahan, 2020). Interestingly, we found the UBL biological process gene mutation status has a very promising predictive value for screening out responders from non-responders both in the POPLAR and OAK cohorts, as well as the MSKCC cohort.

Collectively, this study provided a predictor, UBL biological process gene mutation status, that could be used to distinguish potential responders from non-responders to atezolizumab as second-line therapy among NSCLC patients, with a more promising predictive value in *TP53* mutation-negative subgroups. Furthermore, the predictor UBL biological process also potentially screen the responders from non-responders for the *TP53* mutation-negative NSCLC patients who received other immune checkpoint inhibitors.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Shanghai Chest Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BH, WZ, HW, and JL conceived and designed the experiments. JL, YZ, YL, BY, BZ, MH, YW, YC, ZY, and WZ performed the clinical analysis, bioinformatics analysis, and statistical analysis. JL, YZ, and YL generated the figures and tables. JL wrote the manuscript. BH revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the foundation of Shanghai Chest Hospital (project nos. 2019YNJCM11 and YJXT20190102), the Shanghai Leading Talents Program (2013), the Shanghai Jiao Tong University (project nos. 15ZH4009 and YG2021QN121), the Key Program of Translational Medicine from Shanghai Jiao Tong University School of Medicine (project no. 15ZH1008), the Foundation of Chinese Society of Clinical Oncology (project nos. Y-2019AZZD-0355 and Y-QL2019-0125), and the National Natural Science Foundation of China grants (project no. 31801118).

ACKNOWLEDGMENTS

We thank the patients for their participation in the OAK and POPLAR clinical trials and MSKCC clinical trials, and the investigators for releasing the sequencing data and clinical data.

REFERENCES

Alborelli, I., Leonards, K., Rothschild, S. I., Leuenberger, L. P., Savic Prince, S., Mertz, K. D., et al. (2020). Tumor mutational burden assessed by targeted NGS predicts clinical benefit from immune checkpoint inhibitors in non-small cell lung cancer. *J. Pathol.* 250, 19–29. doi: 10.1002/path.5344

Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., et al. (2018). Comprehensive characterization of cancer driver genes and mutations. *Cell* 173, 371–385. doi: 10.1016/j.cell.2018.02.060

Birkbak, N. J., and McGranahan, N. (2020). Cancer genome evolutionary trajectories in metastasis. *Cancer Cell* 37, 8–19. doi: 10.1016/j.ccell.2019.12.004

Chan, T. A., Yarchoan, M., Jaffee, E., Swanton, C., Quezada, S. A., Stenzinger, A., et al. (2019). Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann. Oncol.* 30, 44–56. doi: 10.1093/annonc/mdy495

Chang, L., Chang, M., Chang, H. M., and Chang, F. (2018). Microsatellite instability: a predictive biomarker for cancer immunotherapy. *Appl. Immunohistochem. Mol. Morphol.* 26, e15–e21. doi: 10.1097/PAI.0000000000000575

Chu, T., Lu, J., Bi, M., Zhang, H., Zhuang, W., Yu, Y., et al. (2021). Equivalent efficacy study of QL1101 and bevacizumab on untreated advanced non-squamous non-small cell lung cancer patients: a phase 3 randomized, double-blind clinical trial. *Cancer Biol. Med.* 18, 816–824. doi: 10.20892/j.issn.2095-3941.2020.0212

Darvin, P., Toor, S. M., Sasidharan Nair, V., and Elkord, E. (2018). Immune checkpoint inhibitors: recent progress and potential biomarkers. *Exp. Mol. Med.* 50, 1–11. doi: 10.1038/s12276-018-0191-1

de Miguel, M., and Calvo, E. (2020). Clinical challenges of immune checkpoint inhibitors. *Cancer Cell* 38, 326–333. doi: 10.1016/j.ccell.2020.07.004

Eguren-Santamaria, I., Sanmamed, M. F., Goldberg, S. B., Kluger, H. M., Idoate, M. A., Lu, B. Y., et al. (2020). PD-1/PD-L1 blockers in NSCLC brain metastases: challenging paradigms and clinical practice. *Clin. Cancer Res.* 26, 4186–4197. doi: 10.1158/1078-0432.CCR-20-0798

Ettinger, D. S., Wood, D. E., Aisner, D. L., Akerley, W., Bauman, J. R., Bharat, A., et al. (2021). NCCN guidelines insights: non-small cell lung cancer, version 2.2021. *J. Natl. Compr. Canc. Netw.* 19, 254–266. doi: 10.6004/jnccn.2021.0013

Fabrizio, D., Cristescu, R., Albacker, L., Snyder, A., Ward, A., Lunceford, J., et al. (2021). Real-world prevalence across 159,872 patients with cancer supports the clinical utility of TMB-H to define metastatic solid tumors for treatment with pembrolizumab. *Ann. Oncol.* 32, 1193–1194. doi: 10.1016/j.annonc.2021.05.805

Fehrenbacher, L., Spira, A., Ballinger, M., Kowanetz, M., Vansteenkiste, J., Mazieres, J., et al. (2016). Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* 387, 1837–1846. doi: 10.1016/S0140-6736(16)00587-0

Gadgeel, S., Rodriguez-Abreu, D., Speranza, G., Esteban, E., Felip, E., Domine, M., et al. (2020). Updated analysis from KEYNOTE-189: pembrolizumab or placebo plus pemetrexed and platinum for previously untreated metastatic nonsquamous non-small-cell lung cancer. *J. Clin. Oncol.* 38, 1505–1517. doi: 10.1200/JCO.19.03136

Gandara, D. R., Paul, S. M., Kowanetz, M., Schleifman, E., Zou, W., Li, Y., et al. (2018). Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat. Med.* 24, 1441–1448. doi: 10.1038/s41591-018-0134-3

Haanen, J., Ernstoff, M., Wang, Y., Menzies, A., Puzanov, I., Grivas, P., et al. (2020). Rechallenge patients with immune checkpoint inhibitors following severe immune-related adverse events: review of the literature and suggested prophylactic strategy. *J. Immunother. Cancer* 8:e000604.

Jamal-Hanjani, M., Wilson, G. A., McGranahan, N., Birkbak, N. J., Watkins, T. B. K., Veeriah, S., et al. (2017). Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* 376, 2109–2121. doi: 10.1056/NEJMoa1616288

Lou, Y., Xu, J., Zhang, Y., Zhang, W., Zhang, X., Gu, P., et al. (2021). Akt kinase LANCL2 functions as a key driver in EGFR-mutant lung adenocarcinoma tumorigenesis. *Cell Death Dis.* 12:170. doi: 10.1038/s41419-021-03439-8

Lu, J., Shi, Q., Zhang, L., Wu, J., Lou, Y., Qian, J., et al. (2019a). Integrated transcriptome analysis reveals KLK5 and L1CAM predict response to anlotinib in NSCLC at 3rd line. *Front. Oncol.* 9:886. doi: 10.3389/fonc.2019.00886

Lu, J., Zhong, H., Chu, T., Zhang, X., Li, R., Sun, J., et al. (2019b). Role of anlotinib-induced CCL2 decrease in anti-angiogenesis and response prediction for nonsmall cell lung cancer therapy. *Eur. Respir. J.* 53:1801562. doi: 10.1183/13993003.01562-2018

Lu, J., Zhong, H., Wu, J., Chu, T., Zhang, L., Li, H., et al. (2019c). Circulating DNA-based sequencing guided anlotinib therapy in non-small cell lung cancer. *Adv. Sci.* 6:1900721. doi: 10.1002/advs.201900721

McGrail, D. J., Pilie, P. G., Rashid, N. U., Voorwerk, L., Slagter, M., Kok, M., et al. (2021). High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. *Ann. Oncol.* 32, 661–672. doi: 10.1016/j.annonc.2021.02.006

Mogi, A., and Kuwano, H. (2011). TP53 mutations in nonsmall cell lung cancer. *J. Biomed. Biotechnol.* 2011:583929. doi: 10.1155/2011/583929

Nagahashi, M., Sato, S., Yuza, K., Shimada, Y., Ichikawa, H., Watanabe, S., et al. (2018). Common driver mutations and smoking history affect tumor mutation burden in lung adenocarcinoma. *J. Surg. Res.* 230, 181–185. doi: 10.1016/j.jss.2018.07.007

Nie, W., Qian, J., Xu, M. D., Gu, K., Qian, F. F., Hu, M. J., et al. (2020). A non-linear association between blood tumor mutation burden and prognosis in NSCLC patients receiving atezolizumab. *Oncoimmunology* 9:1731072. doi: 10.1080/2162402X.2020.1731072

Peters, S., Camidge, D. R., Shaw, A. T., Gadgeel, S., Ahn, J. S., Kim, D. W., et al. (2017). Alectinib versus crizotinib in untreated ALK-positive non-small-cell lung cancer. *N. Engl. J. Med.* 377, 829–838. doi: 10.1056/NEJMoa1704795

Pinheiro, F. D., Teixeira, A. F., de Brito, B. B., da Silva, F. A. F., Santos, M. L. C., and de Melo, F. F. (2020). Immunotherapy–new perspective in lung cancer. *World J. Clin. Oncol.* 11, 250–259. doi: 10.5306/wjco.v11.i5.250

Ramos-Casals, M., Brahmer, J. R., Callahan, M. K., Flores-Chavez, A., Keegan, N., Khamashta, M. A., et al. (2020). Immune-related adverse events of checkpoint inhibitors. *Nat. Rev. Dis. Primers* 6:38. doi: 10.1038/s41572-020-0160-6

Rittmeyer, A., Barlesi, F., Waterkamp, D., Park, K., Ciardiello, F., von Pawel, J., et al. (2017). Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet* 389, 255–265. doi: 10.1016/S0140-6736(16)32517-X

Samstein, R. M., Lee, C. H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., et al. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* 51, 202–206. doi: 10.1038/s41588-018-0312-8

Shankar, B., Zhang, J., Naqash, A. R., Forde, P. M., Feliciano, J. L., Marrone, K. A., et al. (2020). Multisystem immune-related adverse events associated with immune checkpoint inhibitors for treatment of non-small cell lung cancer. *JAMA Oncol.* 6, 1952–1956. doi: 10.1001/jamaoncol.2020.5012

Shaw, A. T., Bauer, T. M., de Marinis, F., Felip, E., Goto, Y., Liu, G., et al. (2020). First-line lorlatinib or crizotinib in advanced alk-positive lung cancer. *N. Engl. J. Med.* 383, 2018–2029. doi: 10.1056/NEJMoa2027187

Zhang, L. L., Lu, J., Liu, R. Q., Hu, M. J., Zhao, Y. M., Tan, S., et al. (2020). Chromatin accessibility analysis reveals that TFAP2A promotes angiogenesis in acquired resistance to anlotinib in lung cancer cells. *Acta Pharmacol. Sin.* 41, 1357–1365. doi: 10.1038/s41401-020-0421-7

Zhao, Y., Liu, J., Cai, X., Pan, Z., Liu, J., Yin, W., et al. (2019). Efficacy and safety of first line treatments for patients with advanced epidermal growth factor receptor mutated, non-small cell lung cancer: systematic review and network meta-analysis. *BMJ* 367:l5460. doi: 10.1136/bmj.l5460

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 September 2021
doi: 10.3389/fgene.2021.707644

Suppression of CPSF6 Enhances Apoptosis Through Alternative Polyadenylation-Mediated Shortening of the *VHL* 3′UTR in Gastric Cancer Cells

Xinglong Shi[1†], Keshuo Ding[2,3†], Qiang Zhao[4], Pengxiao Li[1], Yani Kang[4], Sheng Tan[1] and Jielin Sun[1*]*

[1] Ministry of Education Key Laboratory for Systems Biomedicine, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, China, [2] Department of Pathology, School of Basic Medicine, Anhui Medical University, Hefei, China, [3] Department of Pathology, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [4] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China

OPEN ACCESS

Edited by:
Yuriy L. Orlov,
I.M. Sechenov First Moscow State Medical University, Russia

Reviewed by:
Jian Yang,
Tongji University, China
Chunyan Mou,
Huazhong Agricultural University, China
Eldad Shulman,
Tel Aviv University, Israel

***Correspondence:**
Sheng Tan
tansheng@ustc.edu.cn
Jielin Sun
jlsun@sjtu.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 10 May 2021
Accepted: 12 August 2021
Published: 14 September 2021

Citation:
Shi X, Ding K, Zhao Q, Li P, Kang Y, Tan S and Sun J (2021) Suppression of CPSF6 Enhances Apoptosis Through Alternative Polyadenylation-Mediated Shortening of the VHL 3′UTR in Gastric Cancer Cells. Front. Genet. 12:707644. doi: 10.3389/fgene.2021.707644

Alternative polyadenylation (APA) is an important RNA post-transcriptional process, which can generate diverse mRNA isoforms. Increasing evidence shows that APA is involved in cell self-renewal, development, immunity, and cancer. CPSF6 is one of the core proteins of CFIm complex and can modulate the APA process. Although it has been reported to play oncogenic roles in cancer, the underlying mechanisms remain unclear. The aim of the present study was to characterize CPSF6 in human gastric cancer (GC). We observed that CPSF6 was upregulated in GC. Knockdown of CPSF6 inhibited proliferation and enhanced apoptosis of GC cells both *in vitro* and *in vivo*. Global APA site profiling analysis revealed that knockdown of CPSF6 induced widespread 3′UTR shortening of genes in GC cells, including *VHL*. We also found CPSF6 negatively regulated the expression of VHL through APA and *VHL* short-3′UTR isoform enhanced apoptosis and inhibited cell growth in GC cells. Our data suggested that CPSF6-induced cell proliferation and inhibition of apoptosis were mediated by the preferential usage of poly(A) in *VHL*. Our data provide insights into the function of CPSF6 and may imply potential therapeutic targets against GC.

Keywords: alternative polyadenylation, gastric cancer, apoptosis, CPSF6, VHL

INTRODUCTION

Gastric cancer (GC), one of the top five prevalent cancers, is the third leading cause of cancer mortality across the world, owing to its poor prognosis and diagnoses only at advanced stage with median overall survival less than 1 year (Zhang and Zhang, 2017; Smyth et al., 2020). Many predisposing factors, such as *Helicobacter pylori* infection, age, environmental factors, dietary habits, and so on, assist in the development of GC (Chia and Tan, 2016; Petrovchich and Ford, 2016). Apart from that, genetic alterations are also reported to be involved in GC progression. Abnormal expression of *HER2*, *CDH1*, *TP53*, *FGFR*, *MET*, and other genes are frequently found in different GC types (Oue et al., 2019). The regulation of GC development and progression seems particularly complex; hence, a deeper understanding of the pathogenesis mechanism of GC at the molecular level may be beneficial to identify potential therapeutic targets.

Alternative polyadenylation (APA) is an RNA post-transcriptional process that produces distinct messenger RNA (mRNA) isoforms of a single gene through dictating the length of 3′ untranslated regions (UTRs). Modification of 3′UTR influences mRNA stability, translation, and cellular localization, as 3′UTRs provide major binding sites for protein factors or microRNAs (miRNAs) (Andreassi and Riccio, 2009; Fabian et al., 2010). Recent researches demonstrate that APA can affect various biological functions, including cell self-renewal, differentiation, activation of human immune cells, neuronal activation, and so on (Flavell et al., 2008; Sandberg et al., 2008; Ji and Tian, 2009; Lackford et al., 2014). Of note, APA is also involved in cancer. For example, *Pik3ap1* displayed preferential loss of signal in the proximal 3′UTR in APC and APN cancer subtypes, uniform degradation in LPC cancer subtype, and no change in mature B cells (Singh et al., 2009). Another study also reported that lung (LUSC and LUAD), uterine (UCEC), breast (BRCA), and bladder (BLCA) cancers possessed the highest amount dynamic APA events than the other tumor types, such as kidney renal clear cell carcinoma (KIRC) and head and neck squamous cell carcinoma (Xia et al., 2014).

APA can be determined by over 80 proteins, and the core complex contains four subcomplexes: CPSF, CSTF, CFI, and CFII (Tian and Manley, 2017). Some of them are reported to play an important role in tumorigenesis. It was reported that CSTF2 enhanced oncogenic and metastatic abilities in urothelial carcinoma of the bladder cells (Chen et al., 2018). Intriguingly, NUDT21 (CFIm25) showed a dual role in cancers. For example, NUDT21 was downregulated and functioned as tumor suppressor in glioblastoma, hepatocellular carcinoma, and ovarian cancer (Masamha et al., 2014, 2016; Tan et al., 2018). However, other groups uncovered that knockdown of NUDT21 significantly inhibited cell proliferation and tumorigenicity (Lou et al., 2017; Zhang and Zhang, 2018). These observations implicate sophisticated regulation of APA factors. Meanwhile, how APA takes part in tumor formation is still unclear.

The polyadenylation complex cleavage factor I (CFIm) preferentially binds to UGUA subsequences in pre-mRNAs and controls 3′UTR length, which consists of NUDT21, CPSF6, and CPSF7 (Gruber et al., 2012). In addition, the recombinant NUDT21-CPSF6 subunit complex could also show cleavage activity *in vitro*, which was similar with full CFIm complex (Rüegsegger et al., 1998). As described earlier, NUDT21 regulated tumor growth negatively or positively (Tan et al., 2018). Also, to date, as its partner, the description of CPSF6's function was mainly in HIV infection (Lee et al., 2010; Bejarano et al., 2019). Recent reports also discovered the oncological function of CPSF6, but not in an APA regulatory perspective (Binothman et al., 2017; Zhang et al., 2020). As one of the APA factors, CPSF6 plays an important role in regulation of 3′UTR length. It is reported that knockdown of CPSF6 significantly upregulated the usage of proximal PAS in C2C12 myoblast cells (Li et al., 2015). Besides that, another group also found that knockdown of CPSF6 led to a systematic and transcriptome-wide shift toward proximal poly(A) sites in HEK293 cells (Gruber et al., 2012; Martin et al., 2012). So, it is necessary to discover the roles of CPSF6 in GC progress.

Herein, to expand our knowledge on CPSF6 biological roles, we investigated the phenotypes of CPSF6 knockdown GC cells. We found CPSF6 promoted tumor growth and inhibited apoptosis in GC. Furthermore, we demonstrated CPSF6 negatively regulated von Hippel Lindau (*VHL*), the gene encoding a tumor suppressor, through selective poly(A) usage, thus contributing to tumor growth in GC. Our data provide insights into the function of CPSF6 and may imply potential therapeutic targets.

MATERIALS AND METHODS

Cell Lines and Cell Cultures

Human normal gastric cell line GES1 and GC cancer cell lines (AGS, BGC-823, MKN-28) were purchased from the Chinese Academy of Sciences cell bank (Shanghai, China). All cells were cultured in RPMI-1640 medium (Gibco, United States) supplemented with 10% fetal bovine serum (FBS, Gibco, United States) and penicillin–streptomycin solution (Sangon, China) at 37°C in a 5% CO_2 humidified atmosphere.

Quantitative Real Time RT-PCR (qRT-PCR)

Total RNA was isolated using Trizol (Life Technologies, United States) and reverse transcribed with PrimeScript RT Reagent Kit with gDNA Eraser (Perfect Real Time) (Takara, Japan). Real-time qRT-PCR analysis was performed using TB Green Premix Ex Taq II (Tli RNaseH Plus) (Takara), according to the manufacturer's protocol. The primers used for RT-qPCR analysis are listed as follows: *VHL* 5′-CCCGTATGGCT CAACTTCG-3′ and 5′-GGTTAACCAGAAGCCCATCG-3′; *GAPDH* 5′-CACAGTCCATGCCATCACTG-3′ and 5′-CTTGGCAGCGCCAGTAAG-3′. The expression of genes was normalized to *GAPDH*.

Western Blot Analysis

Cells were lysed in ice-cold RIPA Lysis and Extraction Buffer (Thermo, United States) supplemented with protease (Roche, Switzerland) and phosphatase inhibitor cocktail (Roche). The protein concentrations were determined using BCA Protein Assay Kit (Pierce, United States). Protein lysate was separated by 10% or 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis system (SDS-PAGE) and transferred onto PVDF membranes (Millipore, United States). The membrane was incubated in blocking buffer (1 × TBST with 5% BSA) for 2 h at room temperature and then incubated in diluted primary antibodies: anti-CPSF6 (1:1,000; Abcam, United Kingdom), anti-VHL (1:1,000, Sangon), anti-GAPDH (1:1,000; Proteintech, China), anti-β-actin (1:5,000; Sigma, United States), and anti-β-tubulin (1:1,000; Yeasen, China), and secondary antibody: peroxidase-conjugated goat anti-rabbit IgG (H+L) (1:5,000; Yeasen). The membrane was visualized using ECL start Western Blotting Substrate (GE Healthcare Life Sciences, United States). The intensities of the bands were qualified by ImageJ (National Institutes of Health, United States).

Generation of Stable Cell Lines

The lentiviral vector pLKO.1-puro-shRNA-*CPSF6* was constructed for CPSF6 knockdown as described previously (Tan et al., 2021). The lentiviral vectors pCDH-PGK-*VHL*-CDS-short 3′UTR-SV40-NeoR and pCDH-PGK-*VHL*-CDS-long 3′UTR-SV40-NeoR were constructed for VHL overexpression. 293T cells were transfected for the lentiviral production. The supernatant was harvested after 48 and 72 h, respectively. Then AGS and BGC-823 cells were infected by supernatant for 72 h. Finally, cells were maintained in the presence of puromycin (2 μg/ml) and G418 (600 μg/ml or 1.2 mg/ml) for 2 weeks.

Cell Proliferation and Cell Viability Assays

For cell proliferation assay, cells (1×10^5 cells per well) were plated into 6-well plates. After a certain period of time, count the cell number of fields. Then, calculate the cell number of each well. For cell viability assay, cells (1×10^3 cells per well) were plated into 96-well plates. Five days later, 20 μl of MTT solution (5 mg/ml) was added into each well and cells were cultured for 2 h, and cells were rinsed by PBS and 100 μl DMSO was added into each well. Then the absorbance at 570 nm wavelength was examined.

Colony Formation Assays

GC cells were seeded into 6-well plates (500–1,000 cells/well) and cultured for 2 weeks. Cells were fixed with 4% methanol, stained with crystal violet (0.1%), and photographed.

Cell Apoptosis Assays

The apoptosis rate was evaluated using the Annexin V-FITC/PI Apoptosis Detection kit (Vazyme, China) according to the instructions from the manufacturer. The cells were seeded into 6-well tissue culture plates ($1–4 \times 10^5$ cells/well). After cultured in serum-free medium for 36 h, the cells were collected, washed with ice-cold PBS, and resuspended in 500 μl binding buffer. Then, 5 μl Annexin V-FITC and 5 μl PI were added to the buffer and incubated at room temperature for 10 min in the dark. Cells were analyzed by BD LSRFortessa (BD Biosciences, United States) within 1 h. The TUNEL staining assay was performed using Fluorescein (FITC) Tunel Cell Apoptosis Detection Kit (Servicebio, China) according to the manufacturer's instructions.

Xenograft Model *in vivo*

The design and protocol of *in vivo* experiments were approved by the Institutional Animal Care and Use Committee, Shanghai Jiao Tong University. For tumor growth assay, GC cells were resuspended in PBS/Matrigel Matrix mix at 1:1 ratio. Then 3×10^6 cells in 125 μl solution were injected into the flank of 4-week-old athymic male nude mice. Tumor volumes were calculated as follows: volume = $\frac{1}{2}$ ($L \times W^2$), where L is the length and W is the width. Tumor weight was measured when the mice were executed. The IHC staining of xenograft tumors was performed following the manufacturer's instruction. IHC labeling analysis was reviewed and scored by two independent pathologists.

Dual Luciferase Reporter Assays

The short or the long 3′UTR of the *VHL* gene was cloned into reporter plasmid psiCHECK2 (Promega, United States). GC cells were transfected with the reporter vector in 24-well plates (200 ng/well) using Lipofectamine 3000 (Invitrogen, United States). After 48 h, cells were collected and assayed using dual luciferase assay according to the manufacturer's protocol (Transgen, China).

3T-Seq Analysis

3T-seq analysis was performed as described previously (Lai et al., 2015; Tan et al., 2018). Briefly, 50 μg total RNA was incubated with M280 beads, which were bound by Bio-oligo dT (Lou et al., 2017). After cDNA synthesis, the 3′UTR fragments were released by *Gsu I* digestion and then subjected to deep sequencing. The filtered reads were mapped to UCSC human reference genome (hg19) with bowtie2 (Langmead and Salzberg, 2012). The 3′UTR alteration for each gene in the cell lines was detected by linear trend test (the FDR-adjusted p-value < 0.05) (Fu et al., 2011). These sequence data have been submitted to the EMBL-EBI databases under accession number E-MTAB-9980 and available in the ArrayExpress database.[1]

Statistical Analysis

All experiments were repeated at least three times. The results of this study were analyzed with GraphPad statistical software (GraphPad Software, United States) and shown as mean ± SD. We downloaded the expression data of CPSF6 in normal and GC tissues from The Cancer Genome Atlas (TCGA) database[2] and selected frozen sample data for analysis. To investigate the differential expression of CPSF6 between normal and GC tissues of distinct cancer stages or nodal metastasis status in TCGA, we used the website tool[3].

RESULTS

CPSF6 Is Upregulated in GC

In this study, we first analyzed the CPSF6 levels using IHC assay in GC tissues and the paired non-tumor tissues. IHC staining assay showed that CPSF6 was upregulated in tumor tissues when compared with the surrounding non-tumor counterparts (**Figure 1A**). To explore the clinical relevance of CPSF6 in GC patients, we next compared the expression of CPSF6 mRNA, with the RNA-seq data derived from GC (n = 352) and non-tumor tissues (n = 32) in TCGA. We discovered that the expression of CPSF6 was significantly upregulated in GC tissues (**Figure 1B**). According to 7th edition of the AJCC Cancer Staging Manual, cancer stages and nodal metastasis were recommended to the TNM gastric cancer staging system (Washington, 2010).

[1] http://www.ebi.ac.uk/arrayexpress

[2] https://www.cancer.gov/tcga

[3] http://ualcan.path.uab.edu/analysis.html

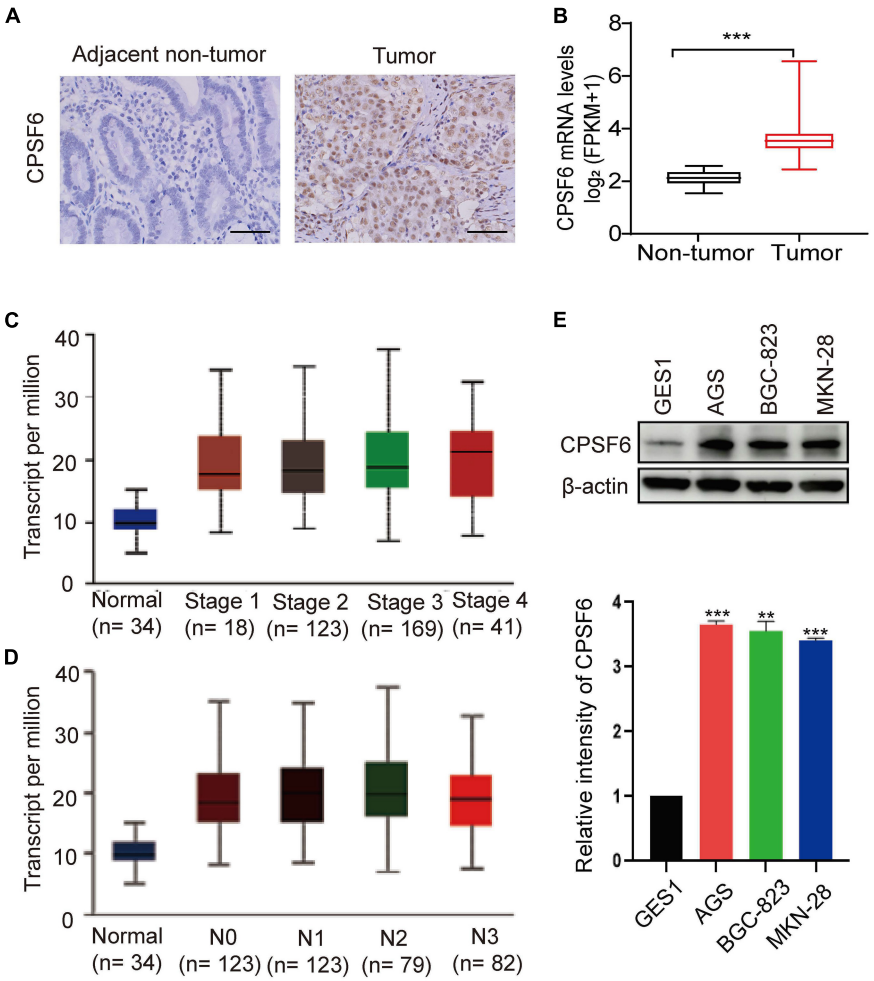


FIGURE 1 | Expression of CPSF6 was upregulated in GC. **(A)** The IHC staining assay of CPSF6 in GC and adjacent tissues of the tumor (scale bar: 100 μm). **(B)** Analysis of CPSF6 mRNA levels of GC (n = 352) and non-tumor tissues (n = 32) in TCGA. **(C)** Expression of CPSF6 in GC based on individual cancer stages in TCGA. **(D)** Expression of CPSF6 in GC based on nodal metastasis status in TCGA. **(E)** Western blot assay for detecting CPSF6 expression in different cell lines. $^{**}0.001 < p < 0.01$; $^{***}p < 0.001$.

To further examine the levels of CPSF6 in distinct GC stages and nodal metastasis, we evaluated the expression of CPSF6 in normal and GC tissues based on individual cancer stages or nodal metastasis. The results showed that the levels of CPSF6 were higher in GC. However, the levels of CPSF6 in different stages and nodal metastasis were similar (**Figures 1C,D** and **Supplementary Tables 1**, **2**). We also analyzed the overall survival curves of CPSF6, and the result showed no significant difference between low and high level of CPSF6 in GC patients (data not shown). The malignancy of GC exhibits the following characteristics: metastasis and uncontrolled proliferation (Hu et al., 2020). The results implied that involvement of CPSF6 in GC formation may be due to the cell growth, rather than the metastasis in GC. We then examined the protein expression of CPSF6 in non-tumorous gastric cell line GES1 and human GC cell lines (AGS, BGC-823, MKN-28). We found that the protein levels of CPSF6 increased in the GC cell lines compared with non-tumorous gastric cell line (**Figure 1E**). Together, our data suggested that CPSF6 was upregulated in GC tissue and cells.

CPSF6 Promotes Cell Growth and Inhibits Apoptosis in GC Cells *in vitro*

To investigate the function of CPSF6 in GC, we knocked down CPSF6 with short hairpin RNAs (shRNAs) in AGS and BGC-823 cells, which have higher levels of CPSF6 (**Figure 1E**). Expression of CPSF6 in these cell lines was obviously repressed (**Figure 2A**). Compared with the control, CPSF6 knockdown in AGS and BGC-823 significantly decreased cell proliferation (**Figure 2B**), viability (**Figure 2C**), and generated less and smaller colonies in colony formation assay (**Figure 2D**). It is reported that the cell cycle and apoptosis affect cell proliferation (Alenzi, 2004). We next investigated if CPSF6 regulated the cell growth through cell cycle and apoptosis in GC cells. The results showed that AGS-shCPSF6 cells increased fraction of S phase cells

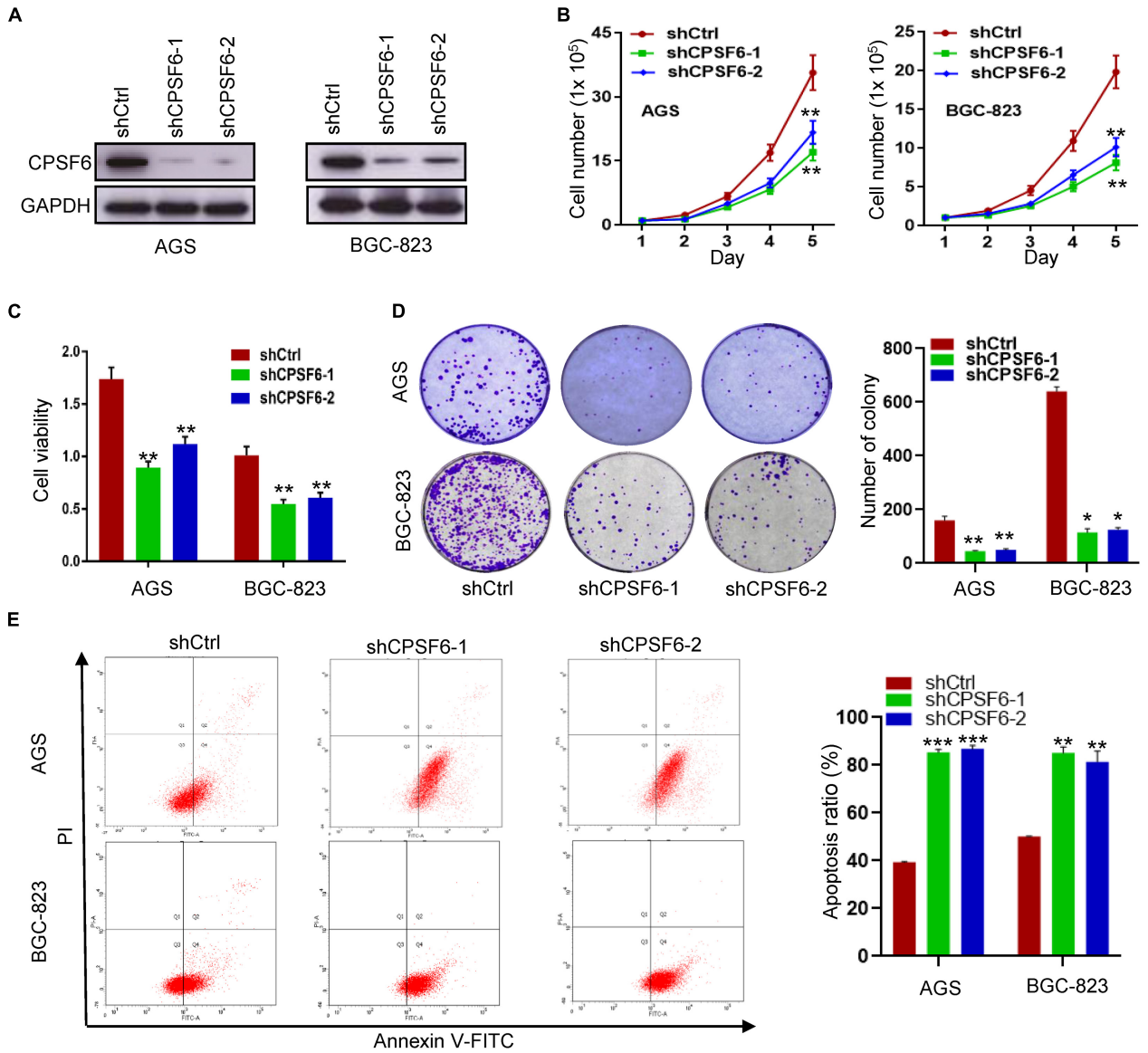


FIGURE 2 | CPSF6 promotes cell growth and inhibits apoptosis in GC cells *in vitro*. **(A)** Western blotting analysis of CPSF6 expression in AGS and BGC-823 cells stably transfected with shRNA plasmids. **(B,C)** Detection of cell proliferation and viability after knockdown of CPSF6 in AGS and BGC-823 cells. **(D)** Detection of colony formation after knockdown of CPSF6 in AGS and BGC-823 cells. **(E)** Detection of apoptosis in AGS and BGC-823 cells. *$0.01 < p < 0.05$; **$0.001 < p < 0.01$; ***$p < 0.001$.

($p < 0.001$), but BGC-823-shCPSF6 cells showed no significant changes ($p > 0.05$) (**Supplementary Figure 1**). Interestingly, both CPSF6 knockdown cell lines exhibited noteworthy increases in cell apoptosis (**Figure 2E**).

CPSF6 Promotes Proliferation and Inhibits Apoptosis of GC Cells *in vivo*

Considering that AGS cells have the highest levels of CPSF6 in the three GC cell lines, the AGS stably transfected cells (AGS-shCtrl, AGS-shCPSF6) were implanted subcutaneously into the flanks of nude mice to further assess the role of CPSF6 *in vivo*. CPSF6 silencing mouse model result also showed similar biological effects in GC cells. We observed a significant decrease in both tumor growth and weight (**Figures 3A,B**) in AGS-shCPSF6 mouse model compared with shCtrl tumors. Besides that, Ki-67 staining result displayed that tumors derived from AGS-shCPSF6 cells exhibited a significantly lower percentage of proliferative cells, as opposed to tumors derived from AGS-shCtrl cells (**Figures 3C,D**). In contrast, TUNEL assay result showed that CPSF6 knockdown increased the apoptosis percentage (**Figure 3E**). Collectively, our results demonstrated that knockdown of CPSF6 inhibited cell growth and increased apoptosis in GC cells.

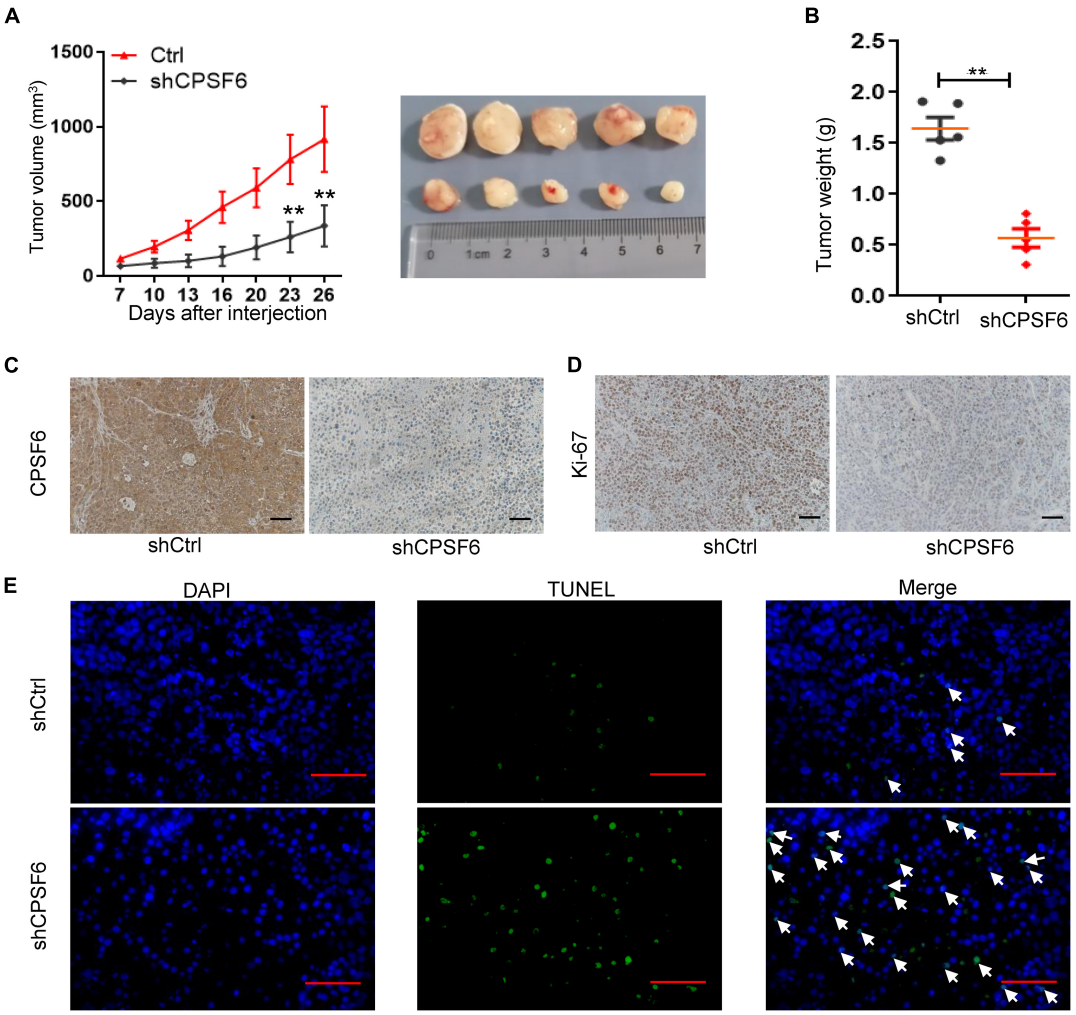


FIGURE 3 | CPSF6 promotes cell growth and inhibits apoptosis in GC cells *in vivo*. **(A,B)** Xenograft nude mouse model of tumorigenesis was used to confirm cell growth and weight after knockdown of CPSF6. **(C,D)** CPSF6 and Ki-67 staining of the respective GC cell–derived tumor sections (scale bar: 100 μm). **(E)** TUNEL assay of the respective GC cell–derived tumor sections (scale bar: 100 μm). The white arrows represent apoptotic cells. (**$0.001 < p < 0.01$).

Knockdown of CPSF6 Modulates the Widespread 3′UTR Shortening in GC Cells

Given that CPSF6 involves in APA formation and its role in promoting cell growth and restraining apoptosis in GC cells as described earlier, we hypothesized that CPSF6 acts as a tumor promoter in GC, at least in part, by influencing APA and 3′UTR. To test this hypothesis, we performed transcriptome-wide APA profiling analysis in CPSF6 knockdown AGS cells and the control cells, using 3T-seq we reported previously (Lai et al., 2015). In total, we gained about 35.7 million reads in AGS-shCtrl cells, as well as 28.9 and 30.9 million reads in AGS-shCPSF6-1 and shCPSF6-2 cells, respectively. About 10.4, 6.6, and 9.1 million reads were uniquely mapped to the reference genome (**Supplementary Table 3**). The internal priming was determined through examining the presence of more than 12 "A"s or continuous 8 "A"s at the region of 1–20 nt downstream the mapping position. After filtering internal priming events, we found majority of qualified reads were mapped to the annotated transcription terminal sites (TTSs) or 3′UTRs (**Figure 4A**), yielding over 10,000 poly(A) sites (**Supplementary Table 3**).

Applying the snowball clustering method (Fu et al., 2011), we found that most of the identified poly(A) sites were located at the 3′ terminus of genes. Particularly, about 60–76% of these poly(A) sites were mapped to TTSs and 52–70% to the 3′UTR regions in each 3T-seq library (**Figure 4B**), according to PolyA_DB database (Wang et al., 2018). More than one third of these poly(A) sites have been annotated in the poly(A) database, and thus the remaining sites may be potential new sites (**Figure 4B**). Also, about 30% of genes harbor three or more poly(A) sites (**Figures 4C,D**). We measured the 3′UTR alteration for CPSF6 responsive genes by introducing cancer 3′UTR length index (CULI) (Fu et al., 2011). A positive CULI suggests that a gene

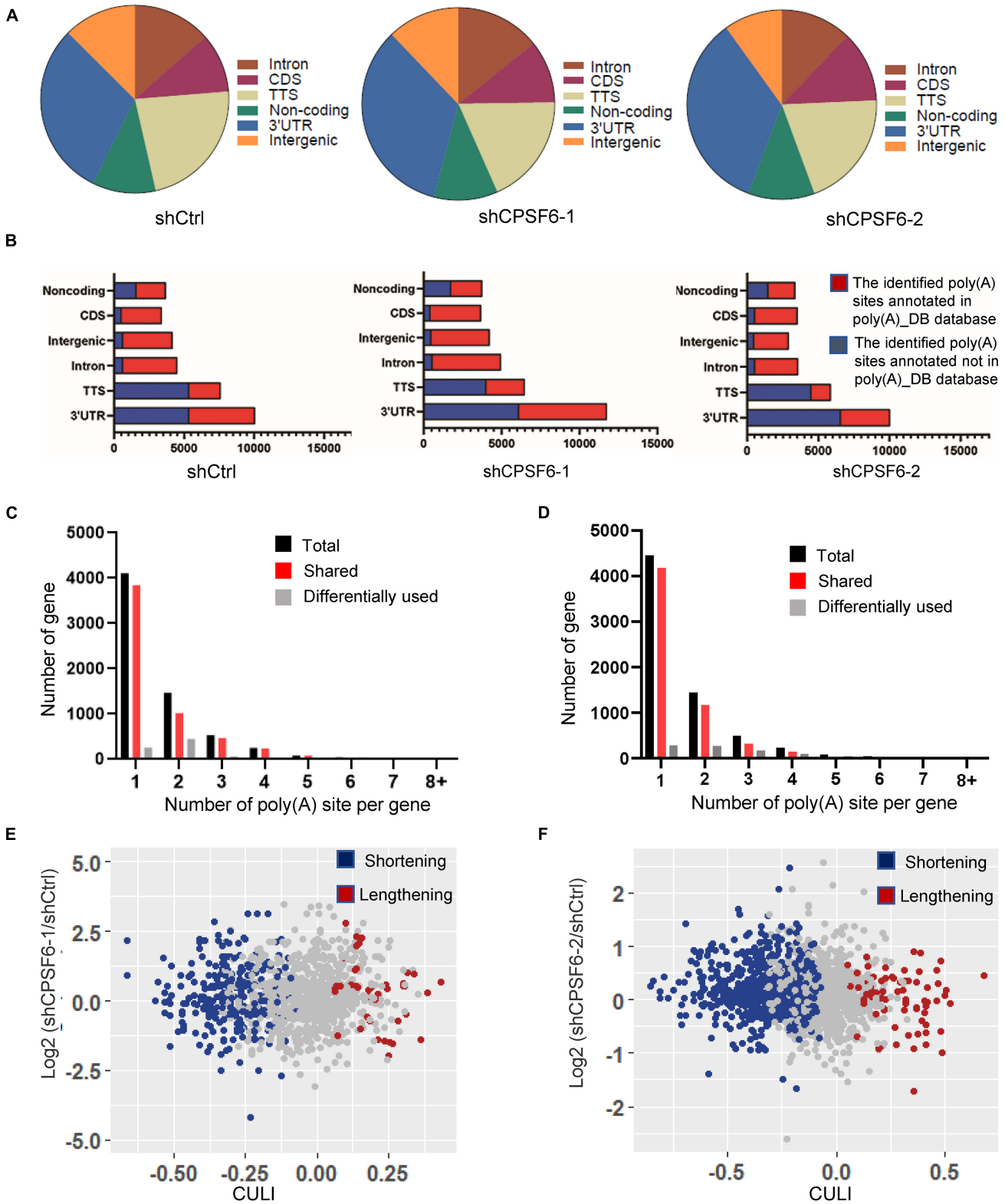


FIGURE 4 | Characteristics of 3T-seq data. **(A)** Genomic locations of 3T-seq reads mapped to the reference genome after filtering internal priming events. **(B)** Genomic distribution of the poly(A) sites. **(C,D)** The statistics of genes with various number of detected poly(A) sites in different libraries. **(E,F)** Scatterplot of CULI in CPSF6 knockdown and control cells [false discovery rate (FDR) < 0.05] in different libraries.

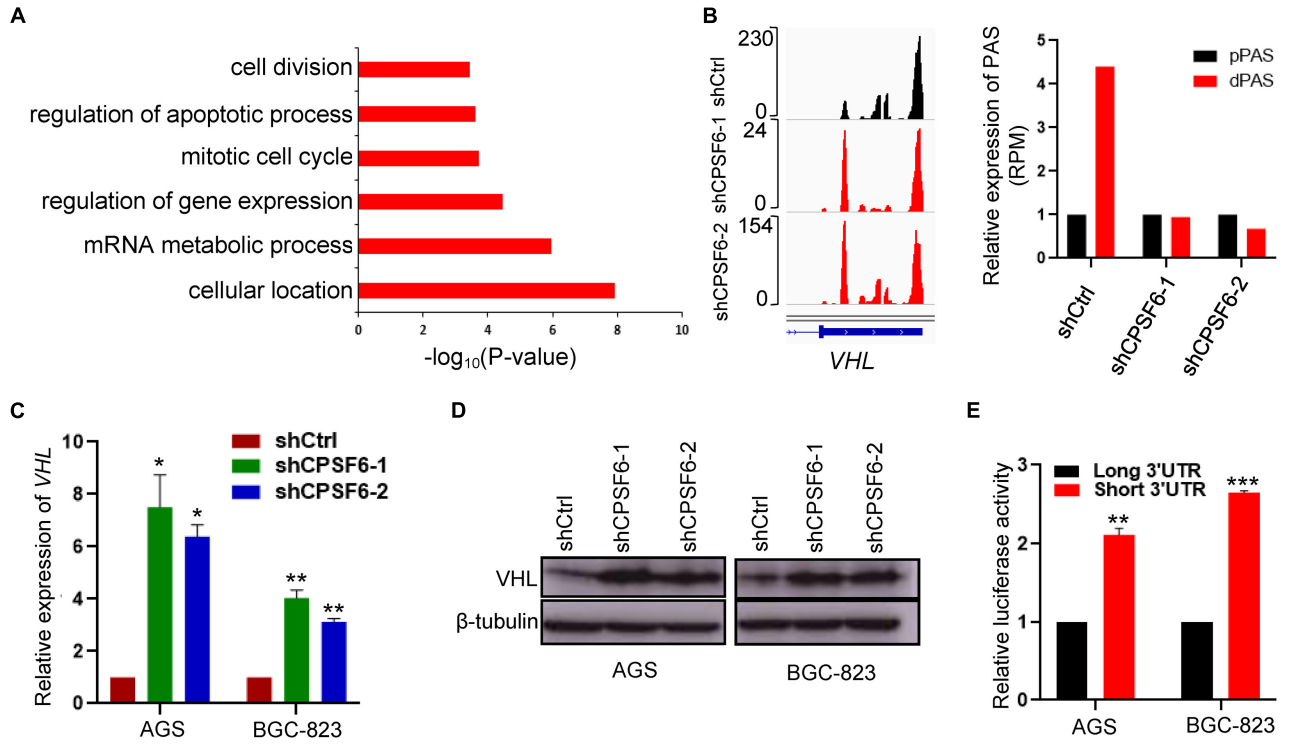


FIGURE 5 | CPSF6 downregulates the expression of VHL. **(A)** GO analysis of genes with shortened 3′UTR after knockdown of CPSF6 in AGS cells. **(B)** CPSF6 induced APA shift of *VHL*. Left, Integrative Genomics Viewer (IGV) genome browser exhibited the poly(A) site usage of *VHL* 3′UTR. Right, histogram showed the relative expression of the isoform with distal polyadenylation site (dPAS) relative to the one with proximal PAS (pPAS). **(C)** The mRNA levels of *VHL* in CPSF6 knockdown AGS and BGC-823 cells. **(D)** The protein levels of VHL in CPSF6 knockdown AGS and BGC-823 cells. **(E)** The luciferase reporter assay for the expression activity of the isoform of *VHL* 3′UTR in AGS and BGC-823 cells. *$0.01 < p < 0.05$; **$0.001 < p < 0.01$; ***$p < 0.001$).

harbors lengthened 3′UTR in CPSF6 knockdown cells compared with the control cells, and a negative CULI indicates the shortened one. Compared with AGS-shCtrl cells, we identified 494 and 633 genes with altered 3′UTR in two independent CPSF6 shRNA knockdown AGS cells, respectively, and 83% and 89% of which displayed a shift from distal to proximal poly(A) site usage and thus possessed shortened 3′UTRs (**Figures 4E,F** and **Supplementary Tables 4**, **5**). This result revealed that knockdown of CPSF6 modulated the widespread 3′UTR shortening in GC cells, which was consistent with previous studies in different cell types (Gruber et al., 2012; Martin et al., 2012).

CPSF6 Downregulates the Expression of VHL Through APA

To ask how CPSF6-modulated APA contributes GC tumorigenesis, we assessed the functional consequence of CPSF6 responsive genes. Thus, to further understand the biological function of APA alternation in GC cells, we collected the 243 genes with shortened 3′UTR, which was simultaneous happened in the two independent AGS CPSF6 knockdown cells and performed the gene ontology (GO) enrichment analysis, using PANTHER classification system (p-value < 0.001, FDR < 0.005)[4] (**Supplementary Table 6**). Notably, GO terms indicated that 3′UTR shortened genes are involved in regulation of apoptosis process and in other biological processes such as cellular and mRNA metabolic processes (**Figure 5A**).

Next, we manually inspected the apoptosis-promoting or proliferation-inhibiting genes under the criterion that the proximal poly(A) site usage increased and the distal poly(A) site usage decreased in both AGS knockdown cells, which were visualized in Integrative Genomics Viewer (IGV) genome browser, yielding three candidates (*IER3IP1*, *IGF1R*, and *VHL*) with top 3′UTR usage fold changes (**Figure 5B** and **Supplementary Figure 2A**). Then, we detected the mRNA expression of the candidate genes by RT-qPCR, and identified *VHL* as CPSF6 responsive gene, which had highest increased mRNA levels in CPSF6 knockdown AGS and BGC-823 cells (**Figure 5C** and **Supplementary Figure 2B**). Moreover, western blotting results indicated that CPSF6 negatively regulated VHL, that is, VHL expression increased in case of CPSF6 knockdown (**Figure 5D**).

A recent study has shown that the gene isoform with short 3′UTR had higher protein expression than the longer one, via escaping the repression of miRNA (Sandberg et al., 2008). As predicted, most of the predicted miRNA binding sites lay outside the proximal poly(A) sites on *VHL* 3′UTR (**Supplementary Figure 2C**). To investigate if the distinct 3′UTR isoforms affect the expression of proteins, we examined

[4]http://geneontology.org

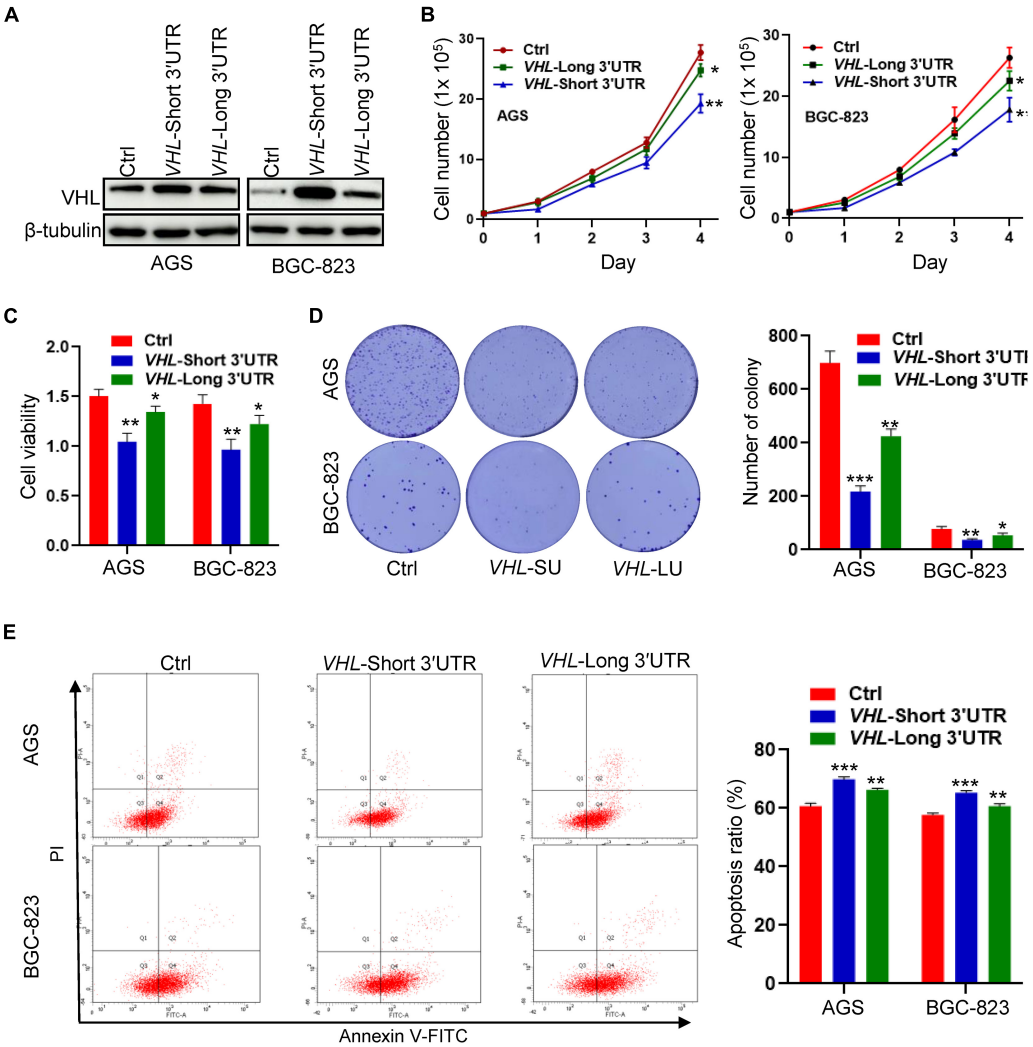


FIGURE 6 | The *VHL* isoform with short 3′UTR inhibits cell growth and promotes apoptosis in GC cells. **(A)** Western blotting analysis of VHL expression in AGS and BGC-823 cells stably transfected with distinct *VHL* isoform with short or long 3′UTR. **(B,C)** Detection of cell proliferation and viability after stably enforced expression of the short- or long-3′UTR isoform of *VHL* in AGS and BGC-823 cells. **(D)** Detection of colony formation after stably enforced expression of the *VHL* isoform with short or long 3′UTR in AGS and BGC-823 cells. **(E)** Apoptosis ratio was analyzed in AGS and BGC-823 cells. *$0.01 < p < 0.05$; **$0.001 < p < 0.01$; ***$p < 0.001$.

expression the short or long *VHL* 3′UTR fragments through the dual luciferase reporter system. As expected, the vector with short 3′UTR showed significantly higher luciferase activities than that with the long 3′UTR, which means short 3′UTR had higher expression ability (**Figure 5E**). Taken together, these data suggested that the expression of VHL was downregulated by CPSF6 through APA in GC.

VHL Short−3′UTR Isoform Enhances Apoptosis and Inhibits Cell Growth in GC Cells

To study the biological function of distinct *VHL* isoforms in GC cells, we stably transfected short- or long−3′UTR isoform of *VHL* into two GC cell lines: AGS and BGC-823. The protein expression of VHL is shown in **Figure 6A**. We found that the short−3′UTR *VHL* isoform dramatically upregulated the protein expression of VHL, as compared with transfection with the long−3′UTR *VHL* isoform. We further observed that of the short−3′UTR isoform of *VHL* in GC cells significantly decreased cell growth, cell viability (**Figures 6B,C**), colony formation (**Figure 6D**), and enhanced cell apoptosis (**Figure 6E**) when compared with that of control cells with the expression of full-length *VHL* and control vector. These data, collectively, suggested high levels of protein produced by *VHL* with shorter 3′UTR may, at least partially, contribute to enhance apoptosis and inhibit cell growth in GC cells.

DISCUSSION

GC is the third leading cause of cancer mortality across the world (Zhang and Zhang, 2017; Smyth et al., 2020). However,

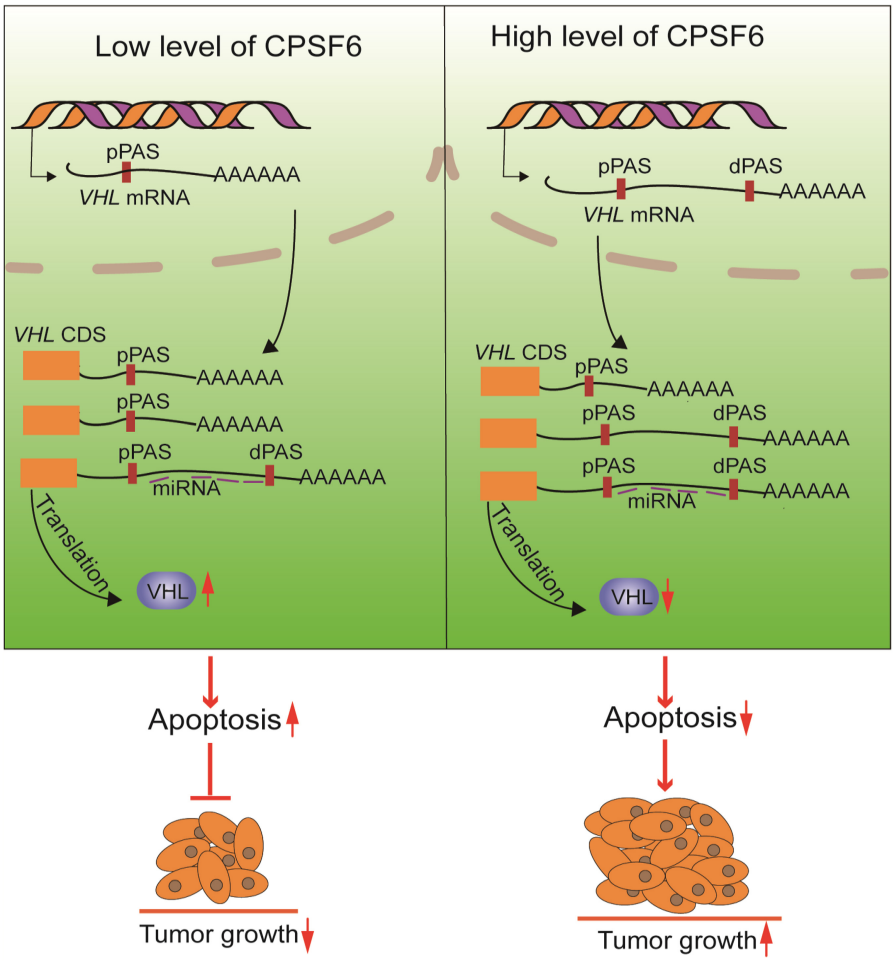


FIGURE 7 | The proposed mechanism of CPSF6 in tumorigenesis of GC progression.

the mechanisms of GC progression are poorly understood. Increasing evidence has proved that APA is involved in many biological functions, including tumorigenesis (Singh et al., 2009; Xia et al., 2014). The APA-mediated regulatory mechanism is determined by numerous APA factors (Tian and Manley, 2017), and some of them are reported to participate in the tumor process. For example, NUDT21 showed a dual role in different cancer types (Masamha et al., 2014; Zhang and Zhang, 2018). This reflects that the regulatory mechanism of APA factors is complex and cancer specific, and it is necessary to characterize the role of APA in different types of cancers. Although CPSF6 played roles in aggressive breast cancer behavior and exhibited a novel CPSF6-RARG fusion variant in acute myeloid leukemia patients (Binothman et al., 2017; Zhang et al., 2020), the role of CPSF6 in GC is still unknown.

In this study, we examined expression of CPSF6 in GC and normal tissues. We found that CPSF6 was upregulated in GC, and what is more, the levels of CPSF6 in different stages and nodal metastasis were similar. We also analyzed the survival curves of CPSF6, and the result showed no significant difference between low and high level of CPSF6 in GC patients (data not shown). This may imply CPSF6 is not involved in metastasis and may contribute to other biological processes (such as proliferation) during GC initiation and progress. Here, we focus on investigating the relationship between CPSF6 and GC growth. Through *in vitro* and *in vivo* assays, we demonstrated CPSF6 enhances the proliferation and inhibits apoptosis in GC.

Given that CPSF6 involves in APA formation and its relevance in GC cells as described previously, we hypothesized that CPSF6 acts as a tumor promoter in GC, at least in part, by influencing APA and 3′UTR. We next performed the APA profiling analysis and identified the majority of APA genes with shortened 3′UTR under the condition of CPSF6 knockdown. The data suggested that knockdown of CPSF6 modulates the widespread 3′UTR shortening in GC cells. In another of our study, we analyzed the usage of 3′UTR in GC relative to normal cells (Lai et al., 2015), and the results showed a global 3′UTR shortening and tendency to use shorter isoforms in GC relative to normal cells. It is very interesting that cancer is associated with global 3′UTR shortening relative to normal cells, and our results seem that CPSF6 promotes cancer via 3′UTR lengthening. Consider that APA factors involve in cancer and regulate the APA progress.

We have analyzed the expression of 20 core APA factors in GC and normal tissues using a website tool (see text footnote 3). Beyond our expectation, we found that 19 of 20 core APA factors were upregulated in GC tissues, and only RBP1 was not regulated significantly ($p = 0.109$) (data not shown). Next, we have searched the function of the 19 APA factors on regulation of 3′UTR. We found that eight factors (CPSF1, CPSF3, CSTF3, CSTF1, PCF11, SYMPK, RBBP6, FIP1L1) prefer to produce the short 3′UTR, and the rest of the factors (WDR33, CPSF4, CPSF2, CSTF2, NUDT21, CPSF6, CPSF7, CLP1, PAPOLA, PABPN1, PABPC1, RBP1) prefer to produce the long 3′UTR (Gruber et al., 2012; Martin et al., 2012; Li et al., 2015). Different factors make different contributions to APA. So, the tendency to use which isoforms in GC is determined by multiple factors. We assumed that although CPSF6 could promote cancer via 3′UTR lengthening, the other APA factors still regulated 3′UTR shortening in GC cells and led GC cells to still prefer to use short 3′UTR.

To understand how CPSF6 exerts its proliferation-promoting and apoptosis-inhibiting function by APA in GC cells, we further characterized CPSF6 regulated genes with shortened 3′UTRs. VHL was demonstrated to contribute to the function of CPSF6-mediated tumor growth in GC cells. We also analyzed the expression of *VHL* in GC relative to normal tissues. The results showed that the levels of *VHL* were higher in GC tissues and the expression of *VHL* in different stages and nodal metastasis was similar (data not shown). As CPSF6 was upregulated in GC cells and downregulated in the *VHL*, how was *VHL* increased in GC tissues? As we all know, many genes are expressed in cancer abnormally, and each of these genes makes partial contribution to cancer. To figure out this question, we analyzed the expression of all genes in GC relative to normal tissues (data not shown). We found that about 3,740 genes were upregulated in GC tissues, including 19 APA factors as described earlier. The 19 APA factors may play an important role in regulating the 3′UTR of *VHL* in different ways and lead to higher expression of *VHL* in GC. Other genes were also increased in GC tissues, for example, *ZNF350* (*ZBRK1*) was upregulated in GC tissues, and it was reported that *ZNF350* could activate *VHL* gene transcription through formation of a complex with *VHL* and p300 in renal cancer (Chen et al., 2015). We also found that about 900 genes were downregulated in GC tissues, including *Daam2*. According to the report, *Daam2* suppressed *VHL* expression and promoted tumorigenesis (Zhu et al., 2017). We supposed that although CPSF6 negatively regulated the *VHL* expression partly, other proteins might increase the level of *VHL* in GC tissues, and which was stronger than the effect of CPSF6 on expression of *VHL*, leading to a higher level of *VHL* in GC tissues.

In this study, we validated that knockdown of CPSF6 leads to a shift from distal to proximal poly(A) site usage, the shortening of 3′UTR in *VHL*, which consequently causes the increased expression of VHL, and inhibition of proliferation in GC cells (**Figure 7**). *VHL* functions as a tumor suppressor gene in some tumors, such as renal cell carcinoma and pheochromocytoma (Richards, 2001), but few studies have reported its role in GC. Interestingly, it was reported that the genetic and epigenetic alterations of the *VHL* were not detected in GC (Cao et al., 2008). Thus, our study may offer a new perspective to understand the regulation of *VHL*. However, how CPSF6 regulates the 3′UTR length and whether other APA factors are involved this progress and the regulation of APA factors in GC and normal tissues still need to be explored. In any case, our data may provide new insights into the understanding of CPSF6's role in GC cells and may imply potential therapeutic targets.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Biomedical Ethics Committee of Anhui Medical University. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Institutional Animal Care and Use Committee, Shanghai Jiao Tong University.

AUTHOR CONTRIBUTIONS

XS, KD, and ST performed most of the experiments and interpretation of the data. QZ, PL, and YK contributed to the analysis of the data and advised on the experimental design. ST and XS wrote the article. ST and JS critically revised the manuscript and contributed to the conception and design. All authors read and approved the final article.

FUNDING

This research was supported by the Shanghai Jiao Tong University Scientific and Technological Innovation Fund (No. 2019TPA09), the Major Projects of Science and Technology Commission of Shanghai (No. 17JC1400800), the Interdisciplinary Program of Shanghai Jiao Tong University (No. YG2019ZDA25), the Special Fund Project for Youth Scientific Research of the Key Laboratory of Systems Biomedicine (Ministry of Education) (No. KLSB2020QN-07), and the Natural Science Foundation of Anhui Province (No. 2008085QH378).

ACKNOWLEDGMENTS

We thank Professor Xiaodong Zhao of Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, China, for the helpful discussion on the conception and design of this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.707644/full#supplementary-material

Supplementary Figure 1 | The cell cycle distribution after knockdown of CPSF6 in AGS and BGC-823 cells.

Supplementary Figure 2 | The candidate genes of CPSF6. **(A)** CPSF6 induced APA shift of candidate genes. UP, Integrative Genomics Viewer (IGV) genome browser exhibited the poly(A) site usage of candidate genes' 3′UTR. Down, histogram showed the relative expression of the isoform with distal polyadenylation site (dPAS) relative to the one with proximal PAS (pPAS). **(B)** The mRNA levels of candidate genes in CPSF6 knockdown AGS cells (up) and BGC-823 cells (down). **(C)** Schematic illustration of the *VHL* isoform with long or short 3′UTR. Positions of the binding sites of miRNA were predicted by website tool (http://www.targetscan.org/) and indicated by purple horizontal lines. CR, CDS region. *$0.01 < p < 0.05$; **$0.001 < p < 0.01$; ***$p < 0.001$.

Supplementary Table 1 | Statistical significance of CPSF6 expression in GC of individual cancer stages.

Supplementary Table 2 | Statistical significance of CPSF6 expression in GC based on nodal metastasis status.

Supplementary Table 3 | The sequencing statistics of AGS 3T-seq libraries.

Supplementary Table 4 | The list of genes with altered 3′UTR in AGS shCPSF6-1 cells.

Supplementary Table 5 | The list of genes with altered 3′UTR in AGS shCPSF6-2 cells.

Supplementary Table 6 | The list of genes with shortened 3′UTR shared by AGS shCPSF6-1 and shCPSF6-2 cells.

REFERENCES

Alenzi, F. Q. (2004). Links between apoptosis, proliferation and the cell cycle. *Br. J. Biomed. Sci.* 61, 99–102. doi: 10.1080/09674845.2004.11732652

Andreassi, C., and Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3 3tability. *Trends Cell Biol.* 19, 465–474. doi: 10.1016/j.tcb.2009.06.001

Bejarano, D. A., Peng, K., Laketa, V., Börner, K., Jost, K. L., Lucic, B., et al. (2019). HIV-1 nuclear import in macrophages is regulated by CPSF6-capsid interactions at the nuclear pore complex. *Elife* 8:e41800. doi: 10.7554/eLife.41800

Binothman, N., Hachim, I. Y., Lebrun, J. J., and Ali, S. (2017). CPSF6 is a Clinically Relevant Breast Cancer Vulnerability Target: role of CPSF6 in Breast Cancer. *EBioMedicine* 21, 65–78. doi: 10.1016/j.ebiom.2017.06.023

Cao, Z., Song, J. H., Kim, C. J., Cho, Y. G., Kim, S. Y., Nam, S. W., et al. (2008). . Genetic and epigenetic analysis of the VHL gene in gastric cancers. *Acta Oncol.* 47, 1551–1556. doi: 10.1080/02841860802001459

Chen, K., Yu, G., Gumireddy, K., Li, A., Yao, W., Gao, L., et al. (2015). ZBRK1, a novel tumor suppressor, activates VHL gene transcription through formation of a complex with VHL and p300 in renal cancer. *Oncotarget* 6, 6959–6976. doi: 10.18632/oncotarget.3134

Chen, X., Zhang, J. X., Luo, J. H., Wu, S., Yuan, G. J., Ma, N. F., et al. (2018). CSTF2-Induced Shortening of the RAC1 3′UTR Promotes the Pathogenesis of Urothelial Carcinoma of the Bladder. *Cancer Res.* 78, 5848–5862. doi: 10.1158/0008-5472.CAN-18-0822

Chia, N. Y., and Tan, P. (2016). Molecular classification of gastric cancer. *Ann. Oncol.* 27, 763–769. doi: 10.1093/annonc/mdw040

Fabian, M. R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA Translation and Stability by microRNAs. *Annu. Rev. Biochem.* 79, 351–379. doi: 10.1146/annurev-biochem-060308-103103

Flavell, S. W., Kim, T.-K., Gray, J. M., Harmin, D. A., Hemberg, M., Hong, E. J., et al. (2008). Genome-Wide Analysis of MEF2 Transcriptional Program Reveals Synaptic Target Genes and Neuronal Activity-Dependent Polyadenylation Site Selection. *Neuron* 60, 1022–1038. doi: 10.1016/j.neuron.2008.11.029

Fu, Y., Sun, Y., Li, Y., Li, J., Rao, X., Chen, C., et al. (2011). Differential genome-wide profiling of tandem 3′ UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res.* 21, 741–747. doi: 10.1101/gr.115295.110

Gruber, A. R., Martin, G., Keller, W., and Zavolan, M. (2012). Cleavage factor Im is a key regulator of 3′ UTR length. *RNA Biol.* 9, 1405–1412. doi: 10.4161/rna.22570

Hu, Y., Zhang, Y., Ding, M., and Xu, R. (2020). LncRNA TMPO-AS1/miR-126-5p/BRCC3 axis accelerates gastric cancer progression and angiogenesis via activating PI3K/Akt/mTOR pathway. *J. Gastroenterol. Hepatol.* 36, 1877–1888. doi: 10.1111/jgh.15362

Ji, Z., and Tian, B. (2009). Reprogramming of 3′ untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One* 4:e8419. doi: 10.1371/journal.pone.0008419

Lackford, B., Yao, C., Charles, G. M., Weng, L., Zheng, X., Choi, E.-A., et al. (2014). Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* 33, 878–889. doi: 10.1002/embj.201386537

Lai, D. P., Tan, S., Kang, Y. N., Wu, J., Ooi, H. S., Chen, J., et al. (2015). Genome-wide profiling of polyadenylation sites reveals a link between selective polyadenylation and cancer metastasis. *Hum. Mol. Genet.* 24, 3410–3417. doi: 10.1093/hmg/ddv089

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lee, K., Ambrose, Z., Martin, T. D., Oztop, I., Mulky, A., Julias, J. G., et al. (2010). Flexible Use of Nuclear Import Pathways by HIV-1. *Cell Host Microbe* 7, 221–233. doi: 10.1016/j.chom.2010.02.007

Li, W., You, B., Hoque, M., Zheng, D., Luo, W., Ji, Z., et al. (2015). Systematic profiling of poly(A)+ transcripts modulated by core 3′ end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet.* 11:e1005166. doi: 10.1371/journal.pgen.1005166

Lou, J. C., Lan, Y. L., Gao, J. X., Ma, B. B., Yang, T., Yuan, Z. B., et al. (2017). Silencing NUDT21 Attenuates the Mesenchymal Identity of Glioblastoma Cells via the NF-κB Pathway. *Front. Mol. Neurosci.* 10:420. doi: 10.3389/fnmol.2017.00420

Martin, G., Gruber Andreas, R., Keller, W., and Zavolan, M. (2012). Genome-wide Analysis of Pre-mRNA 3′ End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3′ UTR Length. *Cell Rep.* 1, 753–763. doi: 10.1016/j.celrep.2012.05.003

Masamha, C. P., Xia, Z., Peart, N., Collum, S., Li, W., Wagner, E. J., et al. (2016). CFIm25 regulates glutaminase alternative terminal exon definition to modulate miR-23 function. *RNA* 22, 830–838. doi: 10.1261/rna.055939.116

Masamha, C. P., Xia, Z., Yang, J., Albrecht, T. R., Li, M., Shyu, A. B., et al. (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature* 510, 412–416. doi: 10.1038/nature13261

Oue, N., Sentani, K., Sakamoto, N., Uraoka, N., and Yasui, W. (2019). Molecular carcinogenesis of gastric cancer: lauren classification, mucin phenotype expression, and cancer stem cells. *Int. J. Clin. Oncol.* 24, 771–778. doi: 10.1007/s10147-019-01443-9

Petrovchich, I., and Ford, J. M. (2016). Genetic predisposition to gastric cancer. *Semin. Oncol.* 43, 554–559. doi: 10.1053/j.seminoncol.2016.08.006

Richards, F. M. (2001). Molecular pathology of von HippelLindau disease and the VHL tumour suppressor gene. *Expert Rev. Mol. Med.* 2001, 1–27. doi: 10.1017/s1462399401002654

Rüegsegger, U., Blank, D., and Keller, W. (1998). Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol. Cell* 1, 243–253. doi: 10.1016/s1097-2765(00)80025-8

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., and Burge, C. B. (2008). Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* 320, 1643–1647. doi: 10.1126/science.1155390

Singh, P., Alley, T. L., Wright, S. M., Kamdar, S., Schott, W., Wilpan, R. Y., et al. (2009). Global changes in processing of mRNA 3′ untranslated regions characterize clinically distinct cancer subtypes. *Cancer Res.* 69, 9422–9430. doi: 10.1158/0008-5472.CAN-09-2236

Smyth, E. C., Nilsson, M., Grabsch, H. I., van Grieken, N. C. T., and Lordick, F. (2020). Gastric cancer. *Lancet* 396, 635–648. doi: 10.1016/S0140-6736(20)31288-5

Tan, S., Li, H., Zhang, W., Shao, Y., Liu, Y., Guan, H., et al. (2018). NUDT21 negatively regulates PSMB2 and CXXC5 by alternative polyadenylation and contributes to hepatocellular carcinoma suppression. *Oncogene* 37, 4887–4900. doi: 10.1038/s41388-018-0280-6

Tan, S., Zhang, M., Shi, X., Ding, K., Zhao, Q., Guo, Q., et al. (2021). CPSF6 links alternative polyadenylation to metabolism adaption in hepatocellular carcinoma progression. *J. Exp. Clin. Cancer Res.* 40:85. doi: 10.1186/s13046-021-01884-z

Tian, B., and Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* 18, 18–30. doi: 10.1038/nrm.2016.116

Wang, R., Nambiar, R., Zheng, D., and Tian, B. (2018). PolyA_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res.* 46, D315–D319. doi: 10.1093/nar/gkx1000

Washington, K. (2010). 7th Edition of the AJCC Cancer Staging Manual: stomach. *Ann. Surg. Oncol.* 17, 3077–3079. doi: 10.1245/s10434-010-1362-z

Xia, Z., Donehower, L. A., Cooper, T. A., Neilson, J. R., Wheeler, D. A., Wagner, E. J., et al. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nat. Commun.* 5:5274. doi: 10.1038/ncomms6274

Zhang, L., and Zhang, W. (2018). Knockdown of NUDT21 inhibits proliferation and promotes apoptosis of human K562 leukemia cells through ERK pathway. *Cancer Manag. Res.* 10, 4311–4323. doi: 10.2147/cmar.S173496

Zhang, X. Y., and Zhang, P. Y. (2017). Gastric cancer: somatic genetics as a guide to therapy. *J. Med. Genet.* 54, 305–312. doi: 10.1136/jmedgenet-2016-104171

Zhang, Z., Jiang, M., Borthakur, G., Luan, S., Huang, X., Tang, G., et al. (2020). Acute myeloid leukemia with a novel CPSF6-RARG variant is sensitive to homoharringtonine and cytarabine chemotherapy. *Am. J. Hematol.* 95, E48–E51. doi: 10.1002/ajh.25689

Zhu, W., Krishna, S., Garcia, C., Lin, C. J., Mitchell, B. D., Scott, K. L., et al. (2017). Daam2 driven degradation of VHL promotes gliomagenesis. *Elife* 6:e31926. doi: 10.7554/eLife.31926

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 20 September 2021
doi: 10.3389/fgene.2021.746114

Pegylated Liposomal Doxorubicin Versus Epirubicin as Adjuvant Therapy for Stage I–III Breast Cancer

Wenxian Hu[1], Kezhen Lv[2], Rongyue Teng[1], Jida Chen[1], Chenpu Xu[1], Lidan Jin[1], Yongxia Chen[1] and Wenhe Zhao[1]*

[1] *Department of Surgical Oncology, Sir Run Run Shaw Hospital, College of Medicine, Zhejiang University, Hangzhou, China,*
[2] *Department of Breast Center, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China*

Background: Conventional anthracyclines, like epirubicin, are cornerstone drugs for breast cancer treatment of all stages, but their cumulative toxicity could cause life-threatening side effects. Pegylated liposomal doxorubicin (PLD), an effective anti-breast cancer drug, has lower toxicity than conventional anthracyclines. This retrospective study compared the efficacy and toxicity profiles between PLD and epirubicin as adjuvant therapy for breast cancer.

Patients and Methods: A total of 1,471 patients diagnosed with stage I–III breast cancer between 2000 and 2018 were included in this study, among which 661 were treated with PLD and 810 with epirubicin, with 45.9 months as the median follow-up time. Anti-breast cancer efficacy was assessed with overall survival (OS) and disease-free survival (DFS), while cardiac toxicity was assessed with left ventricular ejection fraction (LVEF) and electrocardiogram (ECG).

Results: The Kaplan–Meier method and Cox proportional hazards model revealed that there was no statistical difference in OS or DFS between patients treated with PLD and epirubicin, regardless of cancer stages or molecular subtypes (all p-values > 0.05). In addition, patients had significantly better LEVF and ECG data after adjuvant therapy with PLD (both p-values < 0.05).

Conclusion: Based on the large sample size and the long follow-up time of this study, we conclude that PLD has a similar anti-breast cancer efficacy as epirubicin while inducing lower level of cardiac toxicity in Han Chinese. This study suggests that PLD-based adjuvant chemotherapy could be a better option than epirubicin for breast cancer patients especially with existing cardiac disease.

Keywords: breast cancer, pegylated liposomal doxorubicin, epirubicin, adjuvant therapy, efficacy and toxicity

OPEN ACCESS

Edited by:
Hua Li,
Shanghai Jiao Tong University, China

Reviewed by:
Wuguo Deng,
Sun Yat-sen University Cancer Center (SYSUCC), China
Wei Guo,
The Harris Center for Mental health and IDD, United States
Jun Wu,
East China Normal University, China

****Correspondence:***
Wenxian Hu
wenxianhu@zju.edu.cn

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *23 July 2021*
Accepted: *23 August 2021*
Published: *20 September 2021*

Citation:
Hu W, Lv K, Teng R, Chen J, Xu C, Jin L, Chen Y and Zhao W (2021) Pegylated Liposomal Doxorubicin Versus Epirubicin as Adjuvant Therapy for Stage I–III Breast Cancer. Front. Genet. 12:746114. doi: 10.3389/fgene.2021.746114

INTRODUCTION

Breast cancer, the most commonly diagnosed cancer in women, has posed a major threat to women's health globally. In 2018, approximately 2 million women were diagnosed with breast cancer and over 0.6 million women died of breast cancer (Ferlay et al., 2018). Due to the development of better diagnostic tools, therapeutic agents, and surgical techniques, the outcome of breast cancer has been significantly improved in the past decades (Smittenaar et al., 2016). The

use of more powerful chemotherapy agents has markedly increased the survival rate of breast cancer patients, but the resulting toxicities and side effects have become a serious burden. For example, anthracyclines-based chemotherapy could provoke certain acute side effects such as vomiting, nausea and, more severely, congestive cardiac failure (Hortobagyi and Buzdar, 1993; Ryberg et al., 2008).

Doxorubicin and epirubicin are two cornerstone anthracycline drugs used to treat breast cancer of all stages (Khasraw et al., 2012). In spite of its anticancer effects, doxorubicin has showed high toxicity in a large number of studies, and cumulative use of this drug could lead to irreversible cardiomyopathy or liver damage (Mitra et al., 2001; Greish et al., 2004). Epirubicin (a 4′-epimer of doxorubicin) has a similar efficacy as doxorubicin in breast cancer treatment while its dose-dense clinical trials reveal safety comparable to that of doxorubicin (Dang et al., 2008; Nieto et al., 2010). These two drugs have noticeable difference between their toxicity profiles (particularly with respect to cardiotoxicity), and accumulated evidence shows that the use of epirubicin generally brings less side effects compared with doxorubicin, which leads to gradually decreased use of doxorubicin in recent years especially in China (Perez et al., 1991; Khasraw et al., 2012).

Pegylated liposomal doxorubicin (PLD) is formed by incorporating doxorubicin into polyethylene glycol-coated liposomes (Duggan and Keating, 2011). Pegylated liposomal doxorubicin reduces the plasma levels of free doxorubicin and alleviates the toxicity to healthy tissues while maintaining antitumor effects of doxorubicin (Gabizon et al., 2008). Previous studies have shown that PLD has a much longer half-life than that of doxorubicin (73.9 h vs. 10 min), which allows increased uptake of PLD liposomes by tumors and enhanced antitumor effects (Gabizon et al., 2016). The efficacy and safety of PLD-based adjuvant therapy has been investigated in a limited scale for breast cancer patients in previous studies (Lien et al., 2014; Zhao et al., 2017).

The aim of this current study is to compare the efficacy and toxicity profiles between PLD and epirubicin as adjuvant therapy for stage I–III breast cancer. Based on the large number of patients and their clinical data collected at Sir Run Run Shaw Hospital (affiliated with Zhejiang University, Hangzhou, China), we are able to perform a comprehensive evaluation of PLD and epirubicin in different breast cancer stages and molecular subtypes. We find that PLD and epirubicin bring similar overall survival (OS) and disease-free survival (DFS) in stage I–III breast cancer while the toxicity of PLD is considerably improved compared with epirubicin.

PATIENTS AND METHODS

Patient Population

In this retrospective study, a total of 1,471 patients with breast cancer of stages I, II, and III were enrolled at the Department of Surgical Oncology, Sir Run Run Shaw Hospital (affiliated with Zhejiang University, Hangzhou, China). The characteristics and clinical information of patients were retrospectively reviewed and stored in a regularly maintained electronic database for future reference. Among these patients, 661 diagnosed with breast cancer between 2000 and 2018 were treated with PLD as adjuvant chemotherapy; 810 diagnosed between 2001 and 2018 were treated with epirubicin as adjuvant chemotherapy. All patients had ethnicity of Han Chinese and a median follow-up time of 45.9 months. This study was approved by Ethics Committee of Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University.

Patient Treatment

All patients enrolled in this study have not received neoadjuvant chemotherapy. The radical surgery (i.e., the extensive resection of the primary tumor, together with the removal of the surrounding regional lymph nodes) to the primary breast tumor was performed before PLD and epirubicin were administered. PLD was given at 35∼40 mg/m^2 with cyclophosphamide (CTX) every 3 weeks (4 cycles), which could be followed by docetaxel or paclitaxel every 3 weeks (4 cycles). Epirubicin was given at 90∼100 mg/m^2 with the rest of the treatments same as PLDs. Patients were further treated with trastuzumab if the HER2 expression was positive in breast tumor. The above chemotherapy regimens were determined in terms of National Comprehensive Cancer Network[1] guideline for breast cancer. Radiation and endocrine therapy for certain patients were also performed after chemotherapy.

Patient Assessment

The recorded clinical information included date of diagnosis, age at diagnosis, date of recurrence and death, clinical stage, molecular subtype, tumor size, time of surgery, and menopause status. The assessment of tumor was performed with mammography, ultrasound, pulmonary CT, breast MRI, and PET-CT. The clinical stage of tumor was determined based on the American Joint Committee on Cancer (AJCC) 8th Edition.[2] The molecular subtype included HER2-positive, luminal A/B, triple-positive, and triple-negative breast cancers (Cancer Genome Atlas Network, 2012). Toxicity was graded based on the American Society of Clinical Oncology (ASCO) standards.[3] The patients treated with PLD or epirubicin was followed up at a 3-month interval until November 2019 and December 2019, respectively. The health status of patients was assessed at an interval of 3 months with necessary laboratory tests and imaging examinations. During the study, the left ventricular ejection fraction (LVEF) was measured with echocardiogram and electrocardiogram (ECG) was also recorded.

Statistical Analysis

One goal of this study was to compare the efficacy between epirubicin-based and PLD-based chemotherapies, including the difference of OS and DFS. OS was defined as the time from

Abbreviations: PLD, pegylated liposomal doxorubicin; OS, overall survival; DFS, disease-free survival; LVEF, left ventricular ejection fraction; ECG, electrocardiogram; EPI, epirubicin; CI, confidence interval; HR, hazard ratio; CTX, cyclophosphamide; AJCC, American Joint Committee on Cancer.

[1] www.nccn.org/patients

[2] https://cancerstaging.org/

[3] https://www.asco.org/

the start of treatment (i.e., surgery) to the date of death; DFS was defined as the time from the start of treatment to the date of cancer progression. Patients still alive or without disease progressing at the end of the study were censored at the date of last follow-up. OS and DFS were analyzed using the Kaplan–Meier method, and the log-rank test was used to assess the difference between the survival curves. The Cox proportional hazards model was used to calculate the hazards ratios (HR) and corresponding 95% confidence interval (CI) for the two chemotherapies for OS and DFS. Multivariate analysis using the Cox proportional hazards model included other variables of age, menopause status, cancer stage, molecular subtype, and tumor size. Another goal was to compare the toxicity profiles between the epirubicin-based and PLD-based chemotherapies. Student's *t*-test and chi-squared test were used to compare the side effects and the patient characteristics between the two groups. Unless otherwise specified, all statistical tests were two-sided and p-values < 0.05 were considered statistically significant. All statistical analysis was performed with the R language.[4] All figures were prepared with the publicly available R packages (Wickham, 2009; Mailund, 2019; Wang et al., 2020).

RESULTS

Patient Characteristics

Among the 1,471 patients analyzed in this study, 55.1% received epirubicin-based adjuvant chemotherapy and 44.9% received PLD-based adjuvant chemotherapy. For the two groups, the median follow-up time was 63.3 and 30.5 months, respectively. Patients in the two groups have similar menopause status ($p > 0.05$ by chi-squared test; **Table 1**) but slightly different age distributions ($p < 0.05$ by *t*-test, effect size Cohen's $d = -0.16$). Patients receiving epirubicin-based chemotherapy had a higher proportion of stage II breast cancer ($p < 0.05$ by chi-squared test; **Table 1**) and a larger tumor size on average ($p < 0.05$ by KS test), which was adjusted in the multivariate analysis below.

Efficacy of the Two Regimens

The 3–, 5–, and 10-year OS were 98.0%, 94.7%, and 94.7% for PLD-based adjuvant chemotherapy and 98.2%, 96.8%, and 89.5% for epirubicin-based one. The Kaplan–Meier method showed no significant difference of OS between the two groups ($p > 0.05$; **Figure 1A**). Multivariate analysis using the Cox proportional hazards model showed that the HR of the PLD-based regimen was 1.02 (95% CI: 0.50–2.09) after adjusting for multiple factors (age, menopause status, stage, molecular subtype, and tumor size). Meanwhile, the 3–, 5–, and 10-year DFS were 95.5%, 92.6%, and 91.6% for PLD-based adjuvant chemotherapy and 95.0%, 91.4%, and 86.8% for the epirubicin-based one. Similar to OS, the Kaplan–Meier method did not show significant difference of DFS between the two groups ($p > 0.05$; **Figure 1B**) and Cox proportional hazards model yielded an unchanged HR associated with the PLD-based regimen for DFS (HR = 0.94; 95% CI: 0.57–1.56) after adjusting for the above factors.

To assess the efficacy of the two regimens for different stages of breast cancer, we performed survival analysis for patients with stage I, II, and III breast cancer. The 3–, 5–, and 10-year

[4]http://www.r-project.org/

TABLE 1 | Clinical and pathological characteristics of breast cancer patients treated with PLD or epirubicin-based regimens.

Characteristic	Epirubicin (*n* = 810)	PLD (*n* = 661)	Effect size (CI)		*p*-value	Test
Age	48.9 (22.5–73.3)	50.7 (23.5–76.6)	–0.16	Cohen's d	0.003	*t*-test
≤ 50 years	456 (56.3%)	310 (46.9%)				
> 50 years	354 (43.7%)	351 (53.1%)				
Menopause			1.1	Cliff's Delta	0.59	Chi-square test
Yes	38.4% (311)	40.2% (266)				
No	58.0% (470)	57.2% (378)				
Missing	3.6% (29)	2.6% (17)				
Stage			0.17	Cliff's Delta	< 0.001	Chi-square test
I-A	41.1% (333)	56.7% (375)				
II-A	41.2% (334)	30.9% (204)				
II-B	9.5% (77)	5.6% (37)				
III-A	3.3% (27)	3% (20)				
III-B	0.4% (3)	0% (0)				
III-C	0.5% (4)	0.3% (2)				
Missing	4.0% (32)	3.5% (23)				
Molecular subtype			0.1	Cliff's Delta	< 0.001	Chi-square test
HER2-positive	12.3% (100)	7.7% (51)				
Luminal A/B	48.5% (393)	56.0% (370)				
Triple-positive	15.2% (123)	12.6% (83)				
Triple-negative	16.9% (137)	20.6% (136)				
Missing	7.0% (57)	3.2% (21)				
Tumor size	1.9 (0.09–10.7)	1.7 (0.05–11.0)	0.23	Cohen's d	< 0.001	*t*-test

Data are presented with median (range) for age and tumor size, and n (%) for the rest. PLD, pegylated liposomal doxorubicin.

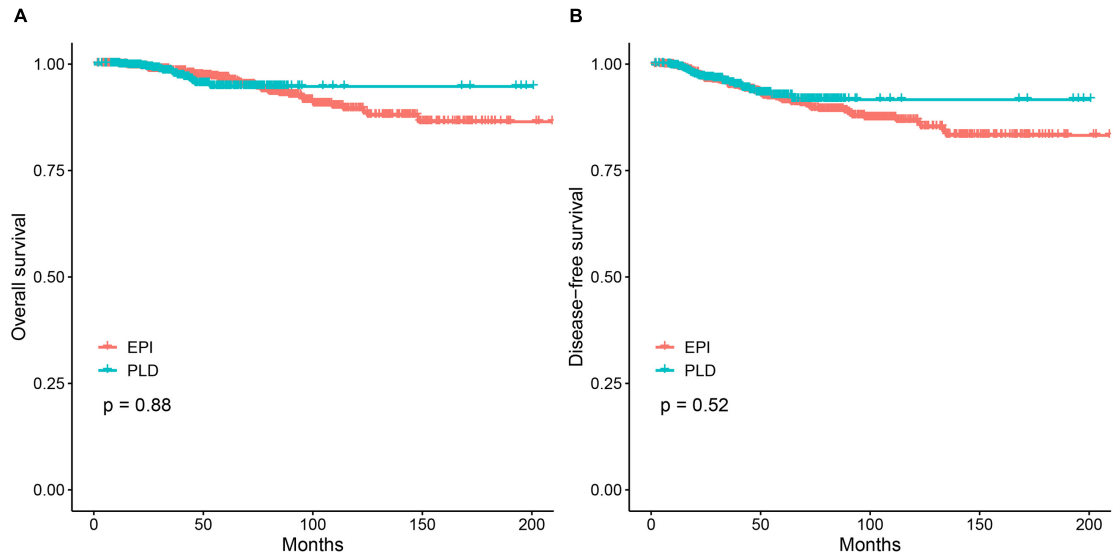


FIGURE 1 | The Kaplan–Meier analysis for the PLD-based regimen vs. the epirubicin-based regimen. **(A)** Comparison of OS between the two regimens. **(B)** Comparison of DFS between the two regimens. PLD, pegylated liposomal doxorubicin; EPI, epirubicin.

OS of the two regimens for different cancer stages is shown in **Supplementary Table 1**. The Kaplan–Meier method showed no significant difference of OS between the two regimens for each stage (**Figure 2A**). The Cox proportional hazards model showed that the HR of the PLD-based regimen was 0.90 (95% CI: 0.38–3.00), 1.37 (95% CI: 0.59–3.15), and 1.17 (95% CI: 0.19–7.08), respectively, for stage I, II, and III breast cancers. The 3–, 5–, and 10-year DFS of the two regimens for different stages is also shown in **Supplementary Table 1**. The Kaplan–Meier method showed no significant difference of DFS between the two regimens for each stage (**Figure 2B**). For DFS, the Cox proportional hazards model showed that the HRs of the PLD-based regimen was 0.74 (95% CI: 0.31–1.76), 1.11 (95% CI: 0.61–2.05), and 1.49 (95% CI: 0.45–4.90) respectively, for stage I, II, and III breast cancers.

We also assessed the efficacy of the two regimens for patients with molecular subtypes of HER2-positive, luminal, triple-positive, and triple-negative. For HER2-positive patients, the Kaplan–Meier method showed no significant difference in OS and DFS (**Figure 3A**), and the Cox proportional hazards model showed that the HR of the PLD-based regimen was 0.83 (95% CI: 0.09–7.40) and 1.25 (95% CI: 0.23–6.82) for OS and DFS, respectively. Similarly, for luminal patients, no difference in OS and DFS was observed (**Figure 3B**) and the HR of PLD-based regimen was 0.67 (95% CI: 0.25–1.83) and 0.63 (95% CI: 0.32–1.26) for OS and DFS, respectively; for triple-positive patients, no difference in OS and DFS was observed (**Figure 3D**) and the HR of the PLD-based regimen was 0.77 (95% CI: 0.08–7.38) and 0.47 (95% CI: 0.10–2.18) for OS and DFS, respectively; and for triple-negative patients, no difference in OS and DFS was observed (**Figure 3C**) and the HR of the PLD-based regimen was 3.30 (95% CI: 0.92–11.83) and 2.18 (95% CI: 0.85–5.57) for OS and DFS, respectively.

Side Effects

We further compared the cardiac toxicity between the two groups. Based on available patients' LVEF data (159 patients from the PLD-based group and 132 from the epirubicin-based group), Student's *t*-test showed that LVEF decreased significantly less in the PLD-based group ($p < 0.05$; **Supplementary Table 2**). For available ECG data (626 patients from the PLD-based group and 479 from the epirubicin-based group), chi-squared test also showed that PLD-based group had more normal cases after treatment ($p < 0.05$; **Supplementary Table 2**). All these results suggested that the PLD-based regimen caused less cardiac toxicity.

DISCUSSION

Although anthracyclines have a long history of being used in adjuvant chemotherapy for breast cancer treatment, their cumulative toxicity has led to life-threatening side effects that restrict further clinical applications (Khasraw et al., 2012; Nicolazzi et al., 2018; Cai et al., 2019). For the purpose of reducing the toxicity, PLD has been developed and is now used in treating both early and advanced breast cancer (Gabizon et al., 2008; Zhao et al., 2017). Pegylated liposomal doxorubicin appeared to be an efficacious and relatively safe option as adjuvant therapy (based on 180 patients), as the toxicities associated with the treatment were mostly manageable (Lu et al., 2016). In this study, we compared survival outcomes and cardiac toxicity between PLD- and epirubicin-based adjuvant therapy for stage I–III breast cancer. We showed that PLD had similar antitumor efficacy as epirubicin while inducing lower level of cardiac toxicity, which suggests that PLD-based adjuvant chemotherapy could be a preferred treatment option for breast cancer patients especially with existing cardiac disease.

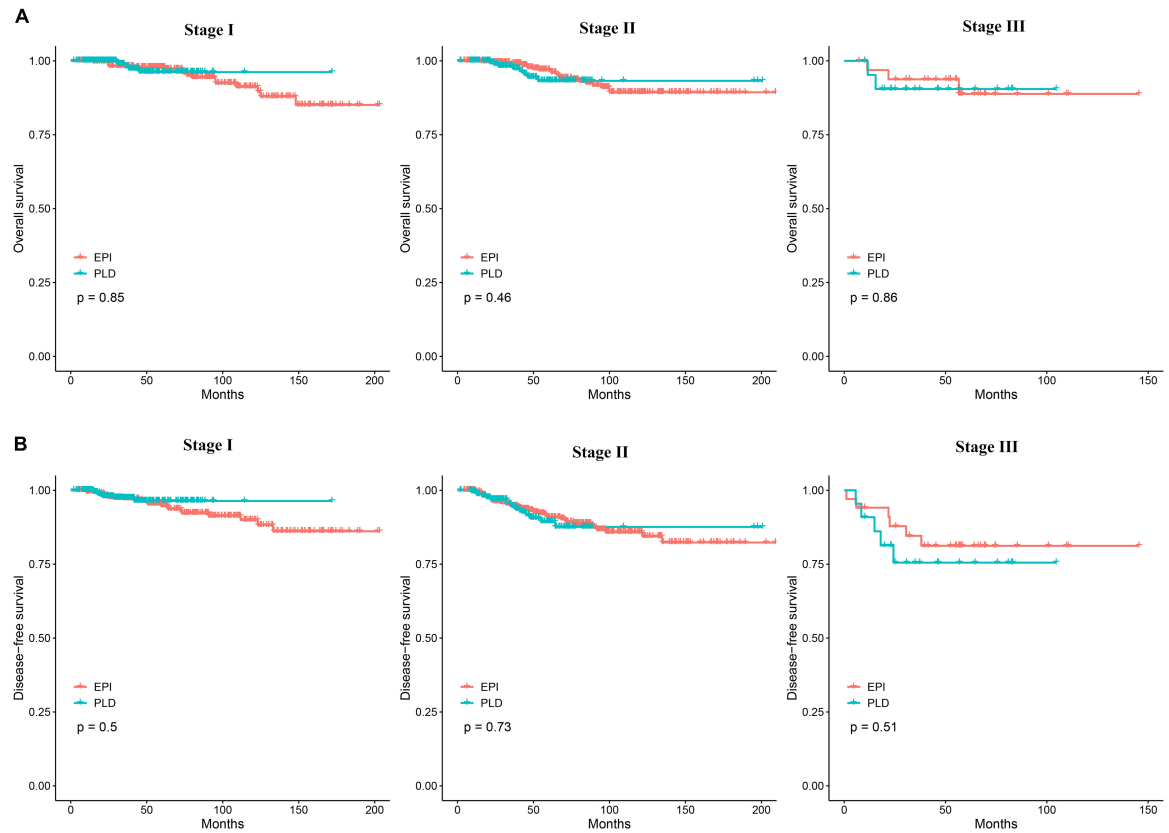


FIGURE 2 | (A) The Kaplan–Meier curves comparing OS between the two groups for stage I, II, and III breast cancer. **(B)** The Kaplan–Meier curves comparing DFS between the two groups for stage I, II, and III breast cancer. OS, overall survival; DFS, disease-free survival.

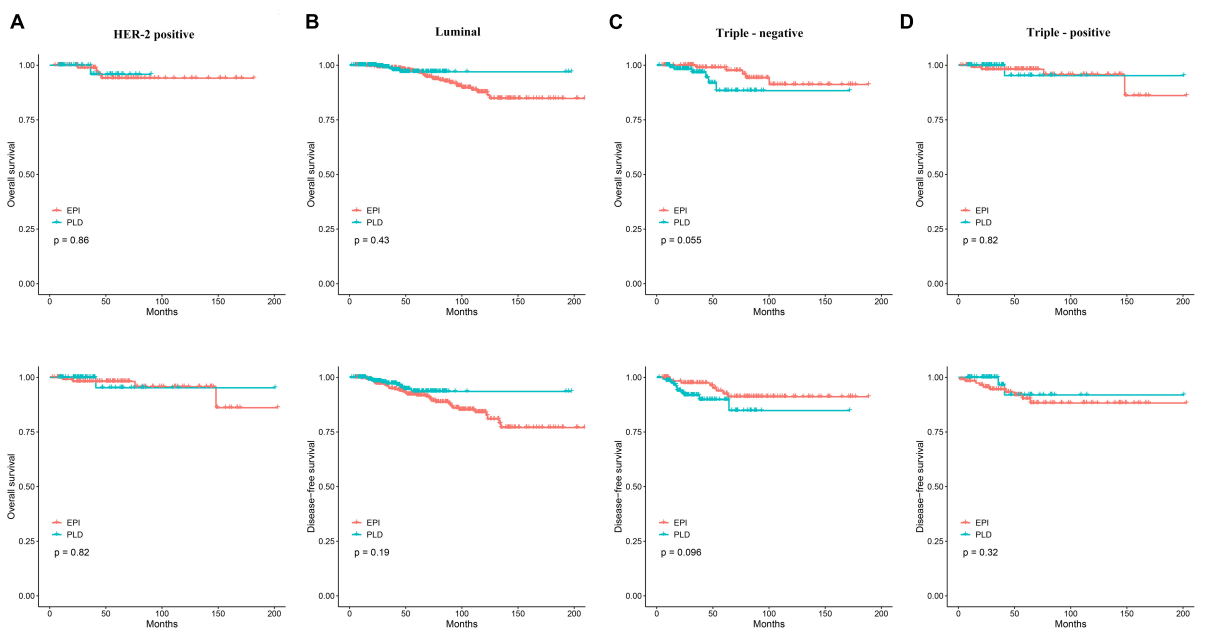


FIGURE 3 | (A) The Kaplan–Meier curves comparing OS and DFS between the two groups in the HER2-positive patients. **(B)** The Kaplan–Meier curves comparing OS and DFS between the two groups in the luminal patients. **(C)** The Kaplan–Meier curves comparing OS and DFS between the two groups in the triple-negative patients. **(D)** The Kaplan–Meier curves comparing OS and DFS between the two groups in the triple-positive patients. OS, overall survival; DFS, disease-free survival.

To the best of our knowledge, this study has at least the following three major strengths. First, this study has a large sample size consisting of 1,471 patients, which allows reliable statistical inference; second, this study has a relatively long follow-up time (up to 209.2 months), which is important for monitoring the cardiac side effects; third, this study has a comprehensive record of clinical information, which makes it possible to compare OS and DFS in patient subgroups (i.e., different stages and molecular subtypes).

One obvious limitation is the retrospective design of this study. Although over 1,400 patients with breast cancer were analyzed in this study, lack of prospective randomization could make our results less convincing. Another limitation is that all patients enrolled in this study (at Sir Run Run Shaw Hospital) were Han Chinese, which makes the conclusions potentially inapplicable to other ethnicity groups even in China. In addition, PLD-based regimen needs to be compared with well-performing non-anthracycline regimens for breast cancer treatment in China (Jones et al., 2006). Nevertheless, this study warrants large-scale prospective studies with more ethnicity groups enrolled in the near future.

CONCLUSION

By comparing the efficacy and toxicity profiles between PLD- and epirubicin-based adjuvant therapy for stage I–III breast cancer, we conclude that, in Han Chinese, PLD has a similar anti-breast cancer efficacy as epirubicin while it induces a lower level of cardiac toxicity. This study also suggests that, for breast cancer patients (especially those with existing cardiac disease), PLD-based adjuvant chemotherapy should be a better option than the epirubicin-based one.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The patient data analyzed in this study are not publicly available due to the confidentiality of the data but are available from the corresponding author on reasonable request. Requests to access these datasets should be directed to WH, wenxianhu@zju.edu.cn.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

WH designed the overall project, analyzed the data, and wrote the manuscript. KL and RT analyzed the data. JC and CX collected clinical information of patients. LJ and YC assisted in data analysis. WZ assisted in manuscript preparation. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Natural Science Foundation of Zhejiang Province (number: LY15H160020).

ACKNOWLEDGMENTS

We thank Chuansheng Hu for assistance in data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.746114/full#supplementary-material

REFERENCES

Cai, F., Luis, M. A. F., Lin, X., Wang, M., Cai, L., Cen, C., et al. (2019). Anthracycline-induced cardiotoxicity in the chemotherapy treatment of breast cancer: preventive strategies and treatment. *Mol. Clin. Oncol.* 11, 15–23.

Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70. doi: 10.1038/nature11412

Dang, C., D'Andrea, G., Lake, D., Sugarman, S., Fornier, M., Moynahan, M. E., et al. (2008). Prolonged dose-dense epirubicin and cyclophosphamide followed by paclitaxel in breast cancer is feasible. *Clin. Breast Cancer* 8, 418–424. doi: 10.3816/cbc.2008.n.050

Duggan, S. T., and Keating, G. M. (2011). Pegylated liposomal doxorubicin a review of its use in metastatic breast cancer. Ovarian cancer, multiple myeloma and AIDS-related kaposi's sarcoma. *Drugs* 71, 2531–2558. doi: 10.2165/11207510-000000000-00000

Ferlay, J., Colombet, M., Soerjomataram, I., Dyba, T., Randi, G., Bettio, M., et al. (2018). Cancer incidence and mortality patterns in Europe: estimates for 40 countries and 25 major cancers in 2018. *Eur. J. Cancer* 103, 356–387.

Gabizon, A. A., Patil, Y., and La-Beck, N. M. (2016). New insights and evolving role of pegylated liposomal doxorubicin in cancer therapy. *Drug Resist. Updat.* 29, 90–106. doi: 10.1016/j.drup.2016.10.003

Gabizon, A., Isacson, R., Rosengarten, O., Tzemach, D., Shmeeda, H., and Sapir, R. (2008). An open-label study to evaluate dose and cycle dependence of the pharmacokinetics of pegylated liposomal doxorubicin. *Cancer Chemother. Pharmacol.* 61, 695–702. doi: 10.1007/s00280-007-0525-5

Greish, K., Sawa, T., Fang, J., Akaike, T., and Maeda, H. (2004). SMA-doxorubicin, a new polymeric micellar drug for effective targeting to solid tumours. *J. Control. Release* 97, 219–230. doi: 10.1016/j.jconrel.2004.03.027

Hortobagyi, G. N., and Buzdar, A. U. (1993). Present status of anthracyclines in the adjuvant treatment of breast cancer. *Drugs* 45 (Suppl 2), 10–19.

Jones, S. E., Savin, M. A., Holmes, F. A., O'Shaughnessy, J. A., Blum, J. L., Vukelja, S., et al. (2006). Phase III trial comparing doxorubicin plus cyclophosphamide with docetaxel plus cyclophosphamide as adjuvant therapy for operable breast cancer. *J. Clin. Oncol.* 24, 5381–5387.

Khasraw, M., Bell, R., and Dang, C. (2012). Epirubicin: is it like doxorubicin in breast cancer? A clinical review. *Breast* 21, 142–149.

Lien, M. Y., Liu, L. C., Wang, H. C., Yeh, M. H., Chen, C. J., Yeh, S. P., et al. (2014). Safety and efficacy of pegylated liposomal doxorubicin-based adjuvant chemotherapy in patients with stage I-III triple-negative breast cancer. *Anticancer. Res.* 34, 7319–7326.

Lu, Y. C., Ou-Yang, F., Hsieh, C. M., Chang, K. J., Chen, D. R., Tu, C. W., et al. (2016). Pegylated liposomal doxorubicin as adjuvant therapy for stage I-III operable breast cancer. *In Vivo* 30, 158–162.

Mailund, T. (2019). *Manipulating Data Frames: dplyr. In:* R Data Science Quick Reference. Berkeley, CA: Apress.

Mitra, S., Gaur, U., Ghosh, P. C., and Maitra, A. N. (2001). Tumour targeted delivery of encapsulated dextran-doxorubicin conjugate using chitosan nanoparticles as carrier. *J. Control. Release* 74, 317–323. doi: 10.1016/s0168-3659(01)00342-x

Nicolazzi, M. A., Carnicelli, A., Fuorlo, M., Scaldaferri, A., Masetti, R., Landolfi, R., et al. (2018). Anthracycline and trastuzumab-induced cardiotoxicity in breast cancer. *Eur. Rev. Med. Pharmacol. Sci.* 22, 2175–2185.

Nieto, Y., Aramendia, J. M., Espinos, J., De la Cruz, S., Fernandez-Hidalgo, O., Santisteban, M., et al. (2010). Sequential administration of dose-dense epirubicin/cyclophosphamide followed by docetaxel/capecitabine for patients with HER2-negative and locally advanced or node-positive breast cancer. *Cancer Chemother. Pharmacol.* 65, 457–465.

Perez, D. J., Harvey, V. J., Robinson, B. A., Atkinson, C. H., Dady, P. J., Kirk, A. R., et al. (1991). A randomized comparison of single-agent doxorubicin and epirubicin as first-line cytotoxic therapy in advanced breast cancer. *J. Clin. Oncol.* 9, 2148–2152.

Ryberg, M., Nielsen, D., Cortese, G., Nielsen, G., Skovsgaard, T., and Andersen, P. K. (2008). New insight into epirubicin cardiac toxicity: competing risks analysis of 1097 breast cancer patients. *J. Natl. Cancer Inst.* 100, 1058–1067. doi: 10.1093/jnci/djn206

Smittenaar, C. R., Petersen, K. A., Stewart, K., and Moitt, N. (2016). Cancer incidence and mortality projections in the UK until 2035. *Br. J. Cancer* 115, 1147–1155.

Wang, Y., Sun, Q., Liang, J., Li, H., Czajkowsky, D. M., and Shao, Z. (2020). Q-Nuc: a bioinformatics pipeline for the quantitative analysis of nucleosomal profiles. *Interdiscip. Sci. Comput. Life Sci.* 12, 69–81. doi: 10.1007/s12539-019-00354-7

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*, (New York, NY: Springer-Verlag).

Zhao, M., Ding, X. F., Shen, J. Y., Zhang, X. P., Ding, X. W., and Xu, B. (2017). Use of liposomal doxorubicin for adjuvant chemotherapy of breast cancer in clinical practice. *J Zhejiang Univ. Sci. B.* 18, 15–26. doi: 10.1631/jzus.b1600303

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 20 September 2021
doi: 10.3389/fgene.2021.722078

Whole-Exome Sequencing on Circulating Tumor Cells Explores Platinum-Drug Resistance Mutations in Advanced Non-small Cell Lung Cancer

Yuanyuan Chang[1,2†], Yin Wang[3†], Boyi Li[3], Xingzhong Lu[3], Ruiru Wang[3], Hui Li[3], Bo Yan[1,2], Aiqin Gu[1], Weimin Wang[1], Aimi Huang[1], Shuangxiu Wu[3] and Rong Li[1,2*]*

[1] Department of Pulmonary Medicine, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [2] Clinical Research Center, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [3] Berry Oncology Corporation, Beijing, China

OPEN ACCESS

Edited by:
Wen-Lian Chen,
Shanghai University of Traditional Chinese Medicine, China

Reviewed by:
Jian Gu,
University of Texas MD Anderson Cancer Center, United States
Zhiyuan Hu,
National Center for Nanoscience and Technology (CAS), China

***Correspondence:**
Shuangxiu Wu
bowusx@aliyun.com
Rong Li
lr22@shchest.org

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 08 June 2021
Accepted: 04 August 2021
Published: 20 September 2021

Citation:
Chang Y, Wang Y, Li B, Lu X, Wang R, Li H, Yan B, Gu A, Wang W, Huang A, Wu S and Li R (2021) Whole-Exome Sequencing on Circulating Tumor Cells Explores Platinum-Drug Resistance Mutations in Advanced Non-small Cell Lung Cancer. Front. Genet. 12:722078. doi: 10.3389/fgene.2021.722078

Circulating tumor cells (CTCs) have important applications in clinical practice on early tumor diagnosis, prognostic prediction, and treatment evaluation. Platinum-based chemotherapy is a fundamental treatment for non-small cell lung cancer (NSCLC) patients who are not suitable for targeted drug therapies. However, most patients progressed after a period of treatment. Therefore, revealing the genetic information contributing to drug resistance and tumor metastasis in CTCs is valuable for treatment adjustment. In this study, we enrolled nine NSCLC patients with platinum-based chemotherapy resistance. For each patient, 10 CTCs were isolated when progression occurred to perform single cell–level whole-exome sequencing (WES). Meanwhile the patients' paired primary-diagnosed formalin-fixed and paraffin-embedded samples and progressive biopsy specimens were also selected to perform WES. Comparisons of distinct mutation profiles between primary and progressive specimens as well as CTCs reflected different evolutionary mechanisms between CTC and lymph node metastasis, embodied in a higher proportion of mutations in CTCs shared with paired progressive lung tumor and hydrothorax specimens (4.4–33.3%) than with progressive lymphatic node samples (0.6–11.8%). Functional annotation showed that CTCs not only harbored cancer-driver gene mutations, including frequent mutations of *EGFR* and *TP53* shared with primary and/or progressive tumors, but also particularly harbored cell cycle–regulated or stem cell–related gene mutations, including *SHKBP1*, *NUMA1*, *ZNF143*, *MUC16*, *ORC1*, *PON1*, *PELP1*, etc., most of which derived from primary tumor samples and played crucial roles in chemo-drug resistance and metastasis for NSCLCs. Thus, detection of genetic information in CTCs is a feasible strategy for studying drug resistance and discovering new drug targets when progressive tumor specimens were unavailable.

Keywords: circulating tumor cells, drug resistance, non-small cell lung cancer, platinum-based chemotherapy, single cell–level WES

INTRODUCTION

Lung cancer is the most common malignancy with the highest incidence and mortality both in China and worldwide (Siegel et al., 2021). Non-small cell lung cancer (NSCLC) comprises about 85% of all lung cancers (LCs), and 50% of NSCLC patients relapse within 5 years after surgery (Taylor et al., 2012; Uramoto and Tanaka, 2014). With the development of detection techniques and targeted therapies, the diagnosis and treatment of NSCLC have made great progress and increased patients' 5-year survival rate in recent years (Carlisle et al., 2020; Siegel et al., 2021). Postoperative adjuvant platinum-based chemotherapy is still the standard regimen for NSCLC patients for whom targeted therapy or progressive after targeted therapy is not applicable. However, a great proportion of NSCLC patients still progress to late-stage disease because of development of drug resistance after treatment. Therefore, it is important to reveal the genetic features of such drug resistance and adjust subsequent treatment to meet precision medicine targets.

Circulating tumor cells (CTCs) are cancer cells detected in the blood of cancer patients. They are assumed to disseminate from the primary tumor lesions and carry the primary lesion's genetic information to relapse tissues through blood circulation during tumor development (Alix-Panabières and Pantel, 2013). CTC's tumorigenic potential is validated in immune-compromised mice by using CTCs from breast cancer, NSCLC, small cell lung cancer (SCLC), and melanoma (Baccelli et al., 2013; Hodgkinson et al., 2014; Girotti et al., 2016; Morrow et al., 2016). It is increasingly accepted to use CTCs in monitoring of early tumor prognosis, recurrence assessment, and therapeutic drug screening in many cancers (Lin et al., 2018). For example, CTCs can be cultured *in vitro* for drug screening (Yu et al., 2014; Carter et al., 2017) or explanted in immune-compromised mice for different therapeutic interventions in SCLC (Hodgkinson et al., 2014; Drapkin et al., 2018) and NSCLC (Morrow et al., 2016; Xu et al., 2018a). Therefore, genetic information in CTCs from drug-resistance patients can help to dissect the mutational evolutionary process driving the transition from a primary tumor to a progressive or metastatic tumor during drug treatment and help us to speculate on the drug-resistance mechanisms underlying tumor progression and explore new drug targets.

In this study, we enrolled a total of nine patients of stage II–IV NSCLC who were all treated with platinum-based chemotherapy independently since diagnosis or after targeted drug treatment, and all progressed within 3 years after chemotherapy (**Table 1** and **Supplementary Table 1**). For each patient at the tumor progressive phase, 10 ml of peripheral blood samples and paired lymph node or lung tumor tissue biopsy or 20 ml of hydrothorax were collected in addition to the paired primary formalin-fixed and paraffin-embedded (FFPE) lung tumor or hydrothorax samples when diagnosed. Then CTCs were isolated from the blood samples using a ClearCell FX1 platform. Whole-exome sequencing (WES) was performed on CTCs and paired primary tumor and progressive samples for each patient to compare the difference of somatic mutation profiles to reveal the genetic information involved in platinum-based chemotherapy and progression of NSCLC.

MATERIALS AND METHODS

Patients and Sampling

Nine relapsed NSCLC patients (Stage II–IV) after treating with platinum-based chemotherapy were enrolled in this study. Their peripheral blood samples, paired tumor biopsies, or hydrothorax samples at disease progressive stage as well as FFPE samples of primary tumor tissues were collected at the Shanghai Chest Hospital of Shanghai Jiao Tong University. Diagnoses of LC were histopathologically confirmed for all patients. Clinical and pathological information are summarized in **Supplementary Table 1**. Informed consent, for sample, acquisition and research purposes in this study were obtained from all patients, and the study was approved by the Medical Ethics Committee of Shanghai Chest Hospital of Shanghai Jiao Tong University to conduct in accordance with the Declaration of Helsinki principles.

Isolation of Circulating Tumor Cells From Blood Samples

A total sample of 10 ml of blood from each cancer patient was collected in a Streck vacutainer tube and processed within

TABLE 1 | Detailed clinical information for each patient.

Patient ID	Gender	Age (y)	Smoking status	Pathology type	Tumor stage[¶]	Chemotherapy drugs	Chemo-therapy cycle
P1	Female	59	Never-smoker	ADC	II/III	Pemetrexed + Carboplatin	4
P2	Male	57	Current smoker	ADC	II/III	Pemetrexed + Carboplatin	8
P3	Male	63	Current smoker	SCC	IV/IV	Gemcitabine + nedaplatin	4
P4	Female	64	Never-smoker	ADC	IV/IV	Pemetrexed + carboplatin	8
P5	Female	66	Never-smoker	ADC	II/IV	Pemetrexed + carboplatin	5
P6	Male	56	Current smoker	ADC	IV/IV	Pemetrexed + carboplatin	4
P7	Female	53	Never-smoker	ADC	III/III	Pemetrexed + carboplatin	4
P8	Male	52	Current smoker	ADC	III/III	Pemetrexed + carboplatin	4
P9	Male	58	Current smoker	SCC	III/IV	Gemcitabine + nedaplatin	4

[¶] *Information before the forward slash (/) refers to the primary-tumor stage, and that after the forward slash (/) refers to the progressive-sample stage. ADC, adenocarcinoma carcinoma; SCC, squamous cell carcinoma; y, years.*

24 h after collection. CTCs were first enriched *via* mass-dependent microfluidics and then manually picked through immunofluorescent staining of both positive staining for pan-cytokeratins (panCK+) and negative staining for CD45 (CD45−) on the basis of an intact Hoechst-stained nucleus (blue fluorescence) as in the method previously reported (Hou et al., 2013; Xu et al., 2018a; Abouleila et al., 2019; Chemi et al., 2019). Anti-CD45 (AB40763, Abcam, United States), anti-panCK (AB7753, Abcam, United States), and Hoechst 33342 were used for imaging of CTCs and white blood cells (WBCs) (**Figure 1**), which were counted under a fluorescence microscope (Olympus IX73), manually picked with a glass capillary into an ice-precooling low-adsorption PCR tube, and stored at −80°C for further experiments. To prevent sequencing failures, biological triplicates to quintuplicates of 10 CTCs were collected from each patient. Biological triplicates of 10 WBCs were collected from the blood sample of P8 and used as the reference control for calling mutations in 10-CTC WES data sets after whole genome amplification (WGA).

Genomic DNA Isolation and Purification From Bulk Tumor Cell Samples

Genomic DNA (gDNA) was extracted from the FFPE samples using the GeneRead DNA FFPE Kit (Qiagen, United States), from the fresh tumor tissue samples using AllPrep DNA/RNA Mini (Qiagen, United States), from bulk WBC samples using the DNA Blood Midi/Mini kit (Qiagen, United States), or large DNA fragments (>500 bp) from the hydrothorax samples using the MagMAX^{TM} Cell-Free DNA Isolation Kit (Life Technology, United States) according to the manufacturer's instructions, respectively. The quality of purified DNA was assayed by gel electrophoresis and quantified by a Qubit® 4.0 Fluorometer (Life Technologies, United States).

Oligo-Cell Whole Genome Amplification

After PBS buffer was removed by centrifugation, 10 CTCs or WBCs were directly lysed in the PCR tube and WGA was performed using the kits of both Preimplantation Genetic Screening of Chromosomal Copy Number Variations (Berry Genomics, China) and/or MALBAC Single Cell DNA Quick-Amp Kit (Yikongenomics, China) according to the manufacturer's protocols. The WGA gDNA was further underwent WES according to the following method.

Library Construction and Whole-Exome Sequencing

The purified gDNA or WGA gDNA was first fragmented into DNA pieces around 300 bp using the enzymatic method (5 × WGS Fragmentation Mix, Qiagen, United States). The WES library was prepared and constructed using the 96 rxn

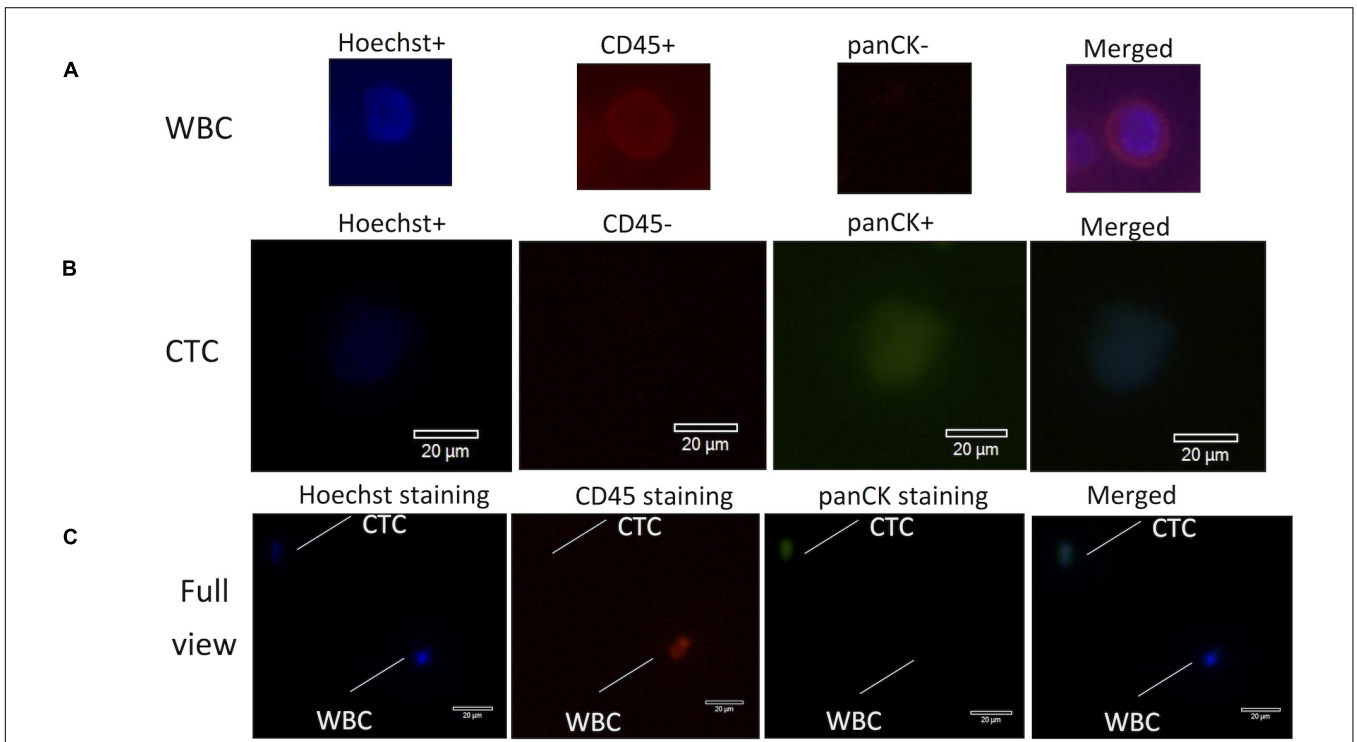


FIGURE 1 | Fluorescence images for identification of CTCs and WBCs stained with antibodies from lung cancer patients. **(A)** WBCs were stained with Hoechst (blue), CD45+ (red), panCK– (dark), and merged images of three-antibody staining. **(B)** CTCs were stained with Hoechst (blue), CD45– (dark), panCK+ (green), and merged images of three-antibody staining. **(C)** Optical images (60× magnification) of isolated CTCs and WBCs under the microscope. Scale bar is 20 mm. CTC morphology is characterized by green fluorescence staining by panCK and nuclear blue staining by Hoechst but negative staining by CD45. CTCs, circulating tumor cells; WBCs, white blood cells.

xGen Exome Research Panel v1.0 (Integrated DNA Technologies, United States), and sequencing was applied on a NovaSeq 6000 platform (Illumina, San Diego, CA, United States) with a 150PE mode according to the manufacturer's protocol (Jia et al., 2020).

Bioinformatics Analysis of Standard Whole-Exome Sequencing Data

For the tumor tissue or hydrothorax samples, single nucleotide variations (SNVs) and small insertions and deletions (Indels) were called and annotated according to previous reports (Jia et al., 2020). Briefly, the non-synonymous SNVs/Indels with VAF > 3% or with VAF > 1% in cancer hot spots as well as with VAF ratio of variant/reference >5 and $p < 0.01$ (Fisher's exact test) for the variant compared between tumor samples and WBC samples were kept for further analysis (**Supplementary Table 3**). Tumor mutation burden (TMB) was defined as the total number of non-synonymous SNVs/Indels per megabase of coding region of a tumor genome in WES.

Bioinformatics Analysis of Whole-Exome Sequencing Data of Circulating Tumor Cells

To obtain high-confidence SNVs/Indels from WES of CTCs, the following filters were applied in addition to the above analysis, and the final results are listed in **Supplementary Table 3**:

(1) Based on annotations using ANNOVAR (Wang et al., 2010), the variants that presented in either the 1000 genomes or the ExAC 03 database were removed.
(2) Mutations that referred to blacklisted genomic regions obtained from the UCSC Genome Table Browser and ENCODE Data Analysis Consortium were removed (Letouzé et al., 2017).
(3) Any variants called in any type of the WBC controls were also excluded.
(4) Variants in CTCs that also presented in either the tumor tissue or malignant pleural effusion samples were kept for further analysis.
(5) After the above filtration, for each patient, if a variant presented in tumor tissues or malignant pleural effusion samples but was not covered in the CTC WES data sets, its read number and sequencing depth were checked in the bam file of CTCs. If the variant reads ≥4 in CTCs and <0.01 (Fisher's exact test) for the variants compared between CTCs and tumor samples, the variant was considered as a true mutation in CTCs and kept for further analysis. If $p > 0.05$, the missing variant in CTCs might be caused by low sequencing coverage in CTC WES, and it was removed from further analysis.

Chemo-Resistance and Stem Cell–Related Mutation Screening

For the genes and variants related to cancer chemotherapy treatment, we downloaded a file, Variant, Gene, and Drug Relationship Data.zip, from PharmGKB annotations[1] as the reference data set and blasted against the variants calling in all the samples to obtain the chemo-resistance genes and mutations list. For the genes and variants related to the function of stem cells, the mutations were screened by the keywords of "stem cell" in the annotation results and then we manually retrieved their influence on cancer stem cells in the published literatures. In addition, all the mutations of each sample were also blasted against the gene lists of 10 canonical oncogenic signaling pathways and 10 DNA damage repair pathways according to the published The Cancer Genome Atlas (TCGA) methods (Knijnenburg et al., 2018; Sanchez-Vega et al., 2018) because these genes are usually cancer-driver genes and contribute to tumor metastasis and drug resistance. These screening results are also listed in **Supplementary Table 3**.

Statistical Analysis and Data Visualization

Statistical analysis used in mutation calling in the CTCs and tumor tissues was described as in the above methods. The Venn and heat map of data visualization were conducted by R/Bioconductor software packages.[2]

RESULTS

Comparison of Whole-Exome Sequencing Quality of Oligo Circulating Tumor Cells, White Blood Cells, and Tumor Specimens

To sequence more CTCs each time as well as to consider the threshold of calling somatic mutations, we pooled 10 CTCs or WBCs together in one tube each time to construct a sequencing library and performed the single cell–level WES as described in the method section. The sequencing data first had filtered out the low-quality reads. More than 83% effective reads were mapped to the reference human genome sequence (hg19/GRCh37) for each sample; the mean coverage of whole exome target regions was 94.8% for CTCs and 87.9% for WBCs as well as 20× depth coverage that accounted for about 87.4 and 61.0% of the whole exome regions for CTCs and WBCs, respectively (**Supplementary Table 2**), similar to the previous report (Chemi et al., 2019). For the WES of primary and progressive tumor or malignant pleural effusion specimens of each patient, at least 93% effective reads were mapped to the reference human genome sequence, and 20× depth reads covered more than 98.3% of the whole exome regions (**Supplementary Table 2**).

Single nucleotide variations/Indels called in CTCs were filtered by those detected in 10 WBCs and bulk WBCs; only SNVs/Indels that were also detected in primary or progressive specimens or hydrothorax samples were involved for further bioinformatics analysis (**Supplementary Table 3**) as described in the method section.

[1] https://www.pharmgkb.org/downloads
[2] http://www.bioconductor.org/

Various Heterogeneity of Somatic Mutations During Tumor Evolution After Platinum-Based Chemotherapy

Whole-exome sequencing data show a distinct somatic mutation spectrum of primary and progressive specimens with platinum-drug resistance features (top 40 genes in **Figure 2A** and **Supplementary Table 3**). The progressive mutational gene functions were significantly enriched to cancer-driver, cell adhesion, and metal ion/calmodulin/ATP binding genes in comparison with those in primary tumors: calcium signaling and cell cycle pathways, protein kinase, transcription factor activity, and calmodulin/ATP/chromatin bindings (**Supplementary Table 4**). Except for patients P3 and P8, who had no paired primary FFPE sample, more than half of patients' (n = 5/7) primary and progressive tumors harbored dozens of overlap mutations, variously accounting for from 16.7 up to 76.9% of total mutations in each patient, indicating various somatic mutation heterogeneity during tumor evolution after platinum-based chemotherapy in different patients, which could also be reflected by TMB changes. Five of seven patients (71.4%) harbored increased TMB (**Figure 2B**) and three (42.9%) had less proportion of shared gene mutations (**Figure 2C**) in the progressive specimens than in the primary ones, indicating that new mutations occurred in the progressive tumors when drug resistance occurred in these patients.

The overlapped mutations between primary and progressive tumor tissues included LC–driver genes (Sanchez-Vega et al., 2018), such as recurrent mutations of *EGFR* and *TP53* among most patients as well as sporadic mutations of essential genes *FAT1*, *NOTCH2*, *NTHL1*, and *POLH* (of P4), *RB1*, *MSH6*, *ATR*, *ATRIP*, and *APTX* (of P6), *SMAD4* (of P7), *CDKN1A*, *RBL1*, *FZD10*, *PER1*, *PPP4R4*, and *FANCM* (of P9) as well as multidrug resistance genes of *MDM2*, *NUP107*, and *UMPS* (of P4) (**Figure 2A** and **Supplementary Table 3**), showing the diversity of somatic mutations during tumor metastasis and platinum-drug resistance among different patients.

Circulating Tumor Cells Harboring Mutations Exclusively From Those in Progressive Lymph Nodes

The sequencing results show 10 CTCs carried dozens of SNVs/Indels on an overall level, and the mutation abundance in CTCs varied in different patients (**Figure 3**). Only one patient's CTC carried more and exclusive SNVs/Indels in comparison with the paired primary and progressive tumor specimens (n = 1/9, P5; **Figure 3F**). Other patients harbored a small number of SNVs/Indels simultaneously identified in CTCs and primary and progressive tumor samples, in total accounting for 2.69% of primary and 2.20% of progressive tumor mutations, respectively, indicating a high level of tumor heterogeneity among CTCs

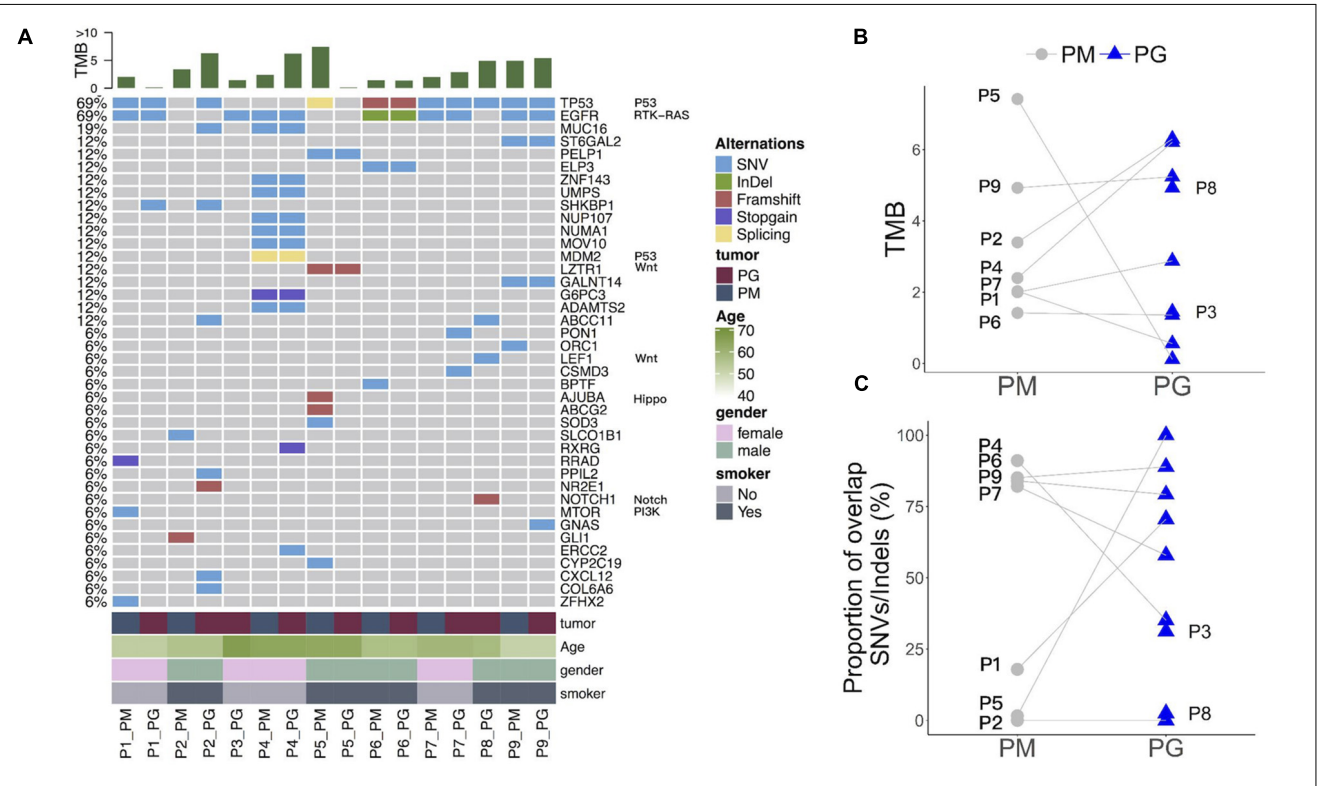


FIGURE 2 | Somatic mutation spectrum and changes of mutation abundance before and after platinum-based chemotherapy treatment of each patient. **(A)** Somatic mutation spectrum in PM and PG. TMB, gene names and related pathways, and clinical information of each patient are on the top, right, and left of the graph. **(B)** Changes of TMB in PM and PG. **(C)** Shared gene mutation proportion in PM and PG. PM, primary specimens; PG, progressive specimens; TMB, tumor mutations burden.

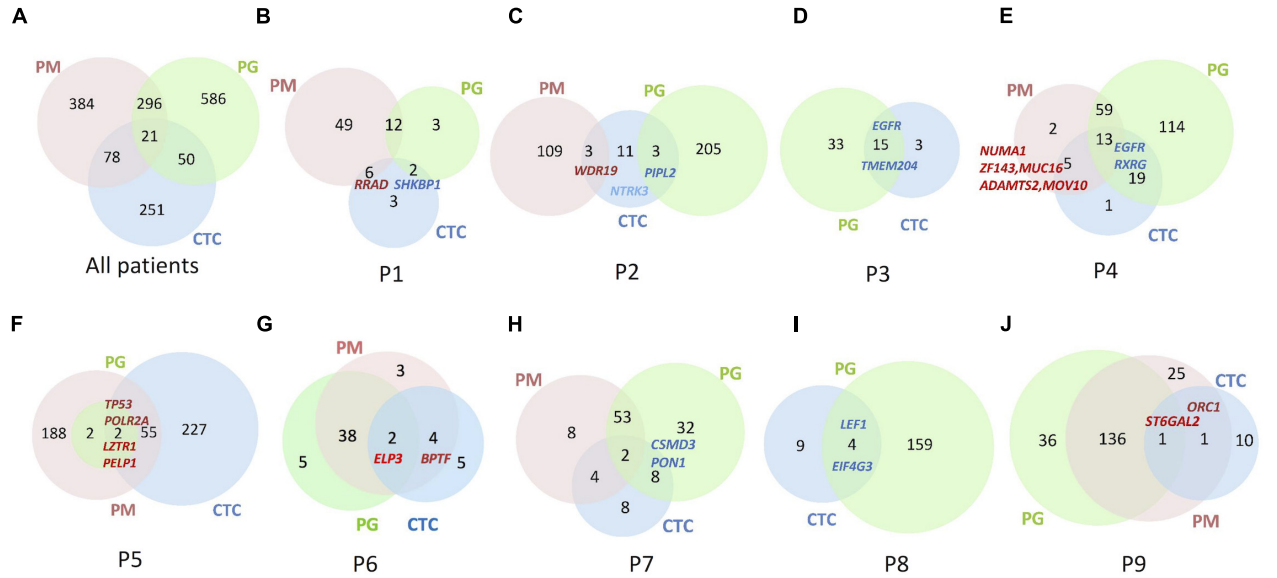


FIGURE 3 | Venn diagrams to show numbers of SNVs/Indels identified in the lung PM, PG, and CTC specimens. **(A)** Summary of all nine patients. **(B–J)** Each patient from Patient 1 to 9. Numbers in Venn diagrams referred to gene numbers. Some essential mutation genes identified are indicated at the corresponding samples. Gene names in bright red denote the mutations shared among PM, PG, and CTC specimens; gene names in dark red denote shared between primary specimens and CTCs; gene names in blue denote shared between progressive specimens and CTC.

and primary and progressive specimens. Notably, the CTCs of three patients (P3, P4, and P6) whose primary and progressive tumor specimens were lung tumor tissue or hydrothorax harbored more mutations shared with progressive samples (33.3% for lung tumor tissue and 4.4–15.6% for hydrothorax) than with progressive lymph nodes (0.6–11.8%). Particularly for P4, the shared mutations included several important tumor metastasis–related and drug-resistance genes, such as *ZNF143*, *MUC16*, *NUMA1*, *ADAMTS2*, and *MOV10* (**Figure 3E** and **Supplementary Table 3**).

Comparing mutations of CTCs with those of primary or progressive samples, except for P2 whose CTCs contained the same numbers of shared mutations between primary and progressive specimens, included a cancer-suspicious gene *WDR19* in primary tumors and a tumor suppressor gene *PPIL2* in progressive lymph nodes. Four patients' CTCs ($n = 4/7$, 57.1%) harbored more mutation overlap with the primary lung tumors, including a nonsense mutation of a tumor-suppressor gene *RRAD* in P1, cancer-driver genes of *POLR2A* and *TP53* in P5, chromatin-remodeling oncogenes *BPTF* and *ELP3* in P6, a stem cell–related gene *ORC1*, and a drug-resistant gene *ST6GAL2* in P9 (**Figure 3** and **Supplementary Table 3**). Two patients' CTCs, including one progressive specimen that was a hydrothorax biopsy ($n = 1/2$, 50.0%) and one progressive specimen that was a lymph node biopsy ($n = 1/6$, 16.7%), harbored mutation profiles more like those of progressive specimens, including chemotherapy-related gene *RXRG* and *EGFR* (with a new mutation site of p.L718Q) in P4 as well as drug-resistance genes *ARHGAP26*, *CSMD3*, and a stem cell–related gene *PON1* in P7. Our findings are consistent with the recent report of different evolutionary mechanisms between lymph node metastasis *via* the lymph system and distance metastasis *via* the blood circulating system (Reiter et al., 2020).

Cell Proliferation and Stem Cell–Related Drug-Resistance Information Detected in CTCs

To interpret drug-resistance genetic information in the CTCs of each patient, the SNVs/Indels of CTCs and progressive and primary specimens were blasted against a cancer chemotherapy database of PharmGKB annotations (see text footnote 1) and a gene list involved in 10 canonical oncogenic signaling pathways and 10 DNA damage repair pathways that were reported to have essential roles in tumor metastasis and drug resistance based on TCGA PanCancer Atlas Project (Knijnenburg et al., 2018; Sanchez-Vega et al., 2018) as described in section "Materials and Methods." The SNVs/Indels were also screened by the keyword "stem cell" in the annotation results and then manually retrieved through public literatures to evaluate their relationship with cancer stem cell features. The results show that the two most frequent mutation genes were *EGFR* and *TP53* (**Figure 4**), which are both cancer-driver and drug-resistance genes in LC. Several cell cycle- and stem cell-related gene mutations were also identified to share between CTCs and primary or progressive specimens depending on different patients (**Table 2**).

For P1, more somatic mutations were detected in the primary tumor (**Figure 3B**), and TMB decreased obviously in the progressive specimens (**Figure 2B**). An *EGFR* activity-regulating gene, *SHKBP1* (SH3KBP1 binding protein 1), mutation was identified in both progressive lymph nodes and CTCs. Its upregulation concomitantly induced activation of AKT through

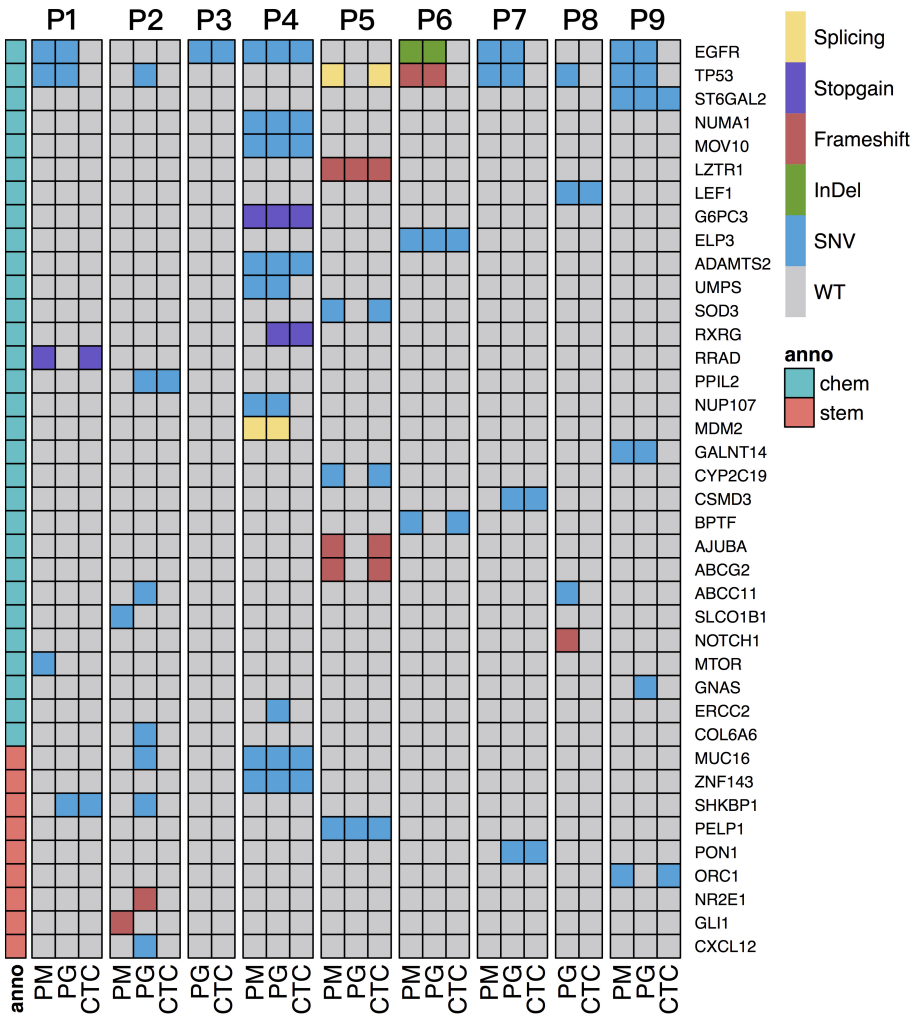


FIGURE 4 | Screening of SNVs/Indels related to chemotherapy and stem cell features shared with PM and/or PG and CTC specimens of all the patients. The gray region represents wild-type. The blank denotes no sequencing reads covered. Sample types are shown on the bottom, and patient IDs are shown on the top of the graph. Mutation genes are listed on the right. The dark blue region represents mutation genes. Chemo denotes chemotherapy-related genes, and stem denotes stem cell feature–related genes, anno denotes annotation, shown on the left column.

the TGFβ pathway and regulation of *EGFR* activity and was highly correlated with epithelial-mesenchymal transition (EMT) and resistance to erlotinib in osteosarcoma cancer stem cell–like cells (Wang et al., 2019). In addition, a nonsense mutation at C terminal (p.R263X) of *RRAD* gene was detected in both the primary tumor and CTCs. It is a tumor-suppressor gene and encodes a RAS-related glycolysis inhibitor and calcium channel regulator (*RRAD*), a small Ras-related GTPase that has been implicated in metabolic disease and several types of cancer. For example, its expression decreased in hepatocellular carcinoma (HCC) tumor tissues and played an important role in regulating aerobic glycolysis and cell invasion and metastasis of HCC (Shang et al., 2016). Therefore, both *SHKBP1* and *RRAD* probably contributed to tumor development and drug resistance through RTK/RAS-related regulations.

For P2, increased TMB and more chemotherapy drug–related gene mutations were detected in the progressive tumor (**Figure 3C**), including *TP53*, *CXCL12*, *MUC16*, *SHKBP1*, *ABCC11*, *CYP2C9*, and a nonsense mutation (p.Y158X) of *CYP26B1*, than those of the primary tumor, such as *SLCO1B1*. However, only mutations of *PPIL2* and *JAKMIP3* were identified in the progressive tumor and CTCs. Upregulation of *PPIL2* was reported to inhibit EMT and tumor invasion by interacting with the classical EMT transcription factor, *SNAI1*, to enhance its ubiquitin-dependent degradation in breast cancer (Jia et al., 2018). However, *JAKMIP3*'s role in LC is not known.

For P3 and P8, who only had progressive tumor and CTC samples (**Figures 3D,I**), the driver genes *EGFR* with p.L858R mutation and *LEF1* of the Wnt signaling pathway were detected, respectively, in both CTCs and progressive tumors of each patient. Mutation of *TAOK1* of the Hippo pathway and *KEAP1* of the NRF2 pathway in the progressive tumors of P3 as well as *NOTCH1* (frameshift deletion of p.R1824fs) of Notch pathway, *IRS1* and *RASAL2* of RTK-RAS pathway, *TP53* of

TABLE 2 | Genetic information identified in CTCs of platinum-drug resistance patients.

Patient ID	Gene name§	Sample (PM/PG)	Sample (CTCs)	Functions and pathways	Drug resistance	References
P1	*SHKBP1*	PG	+	RTK/RAS and TGFβ-related signaling in stem cells	Erlotinib	Wang et al., 2019
	RRAD	PM	+	Ras-related GTPase signaling		Shang et al., 2016
P2	*SHKBP1*	PG	–	RTK/RAS and TGFβ-related signaling in stem cells	Erlotinib	Wang et al., 2019
	PPIL2	PG	+	Epithelial-mesenchymal transition (EMT)		Jia et al., 2018
P3	*EGFR*	PG	+	RTK/RAS signaling	Multi-drug resistance	
P4	*NUMA1*	PM, PG	+	Cell cycle	Multi-drug resistance	Qin et al., 2017
	ZNF143	PM, PG	+	Cell cycle, apoptosis	Multi-drug resistance	Izumi et al., 2010
	MUC16	PM, PG	+	Apoptosis	Cisplatin, gemcitabine	Lakshmanan et al., 2017; Kanwal et al., 2018
P5	*LZTR1*	PM, PG	+	Wnt signaling		Motta et al., 2019
	PELP1	PM, PG	+	Estrogen receptor signaling		Słowikowski et al., 2015; Wang and Bao, 2020
P6	*BPTF*	PM	+	Wnt signaling and TGFβ signaling		Liu et al., 2011
	ELP3	PM, PG	+	Chromatin organization, PI3K/ATK signaling	Multi-drug resistance	Xu et al., 2018b
P7	*CSMD3*	PG	+	Regulation of dendrite development	Etoposide	Qiu et al., 2019
	PON1	PG	+	Regulating stem-cell related *Nanog* expression		☆
P8	*TP53*	PG	+	Cell cycle arrest, apoptosis	Multi-drug resistance	
P9	*ORC1*	PM, PG	+	Cell cycle, stem cell related		Zhang et al., 2020

PM, primary samples; PG, progressive samples; CTC, circulating tumor cells.
+, positive in CTCs; –, negative in CTCs.
§The genes' full names were in the main text.
☆https://www.genecards.org/cgi-bin/carddisp.pl?gene=PON1.

p53/CPF pathway, and a multidrug-resistance gene *ABCC11* in the progressive tumors of P8 were also detected, indicating their roles are involved in the evolution of tumor progression and possibly also are involved in drug resistance.

Increased TMB, more progressive somatic mutations (**Figure 3E**), and more shared drug-resistance or cancer stem cell–related gene mutations were detected in P4 than in other patients (**Figure 4** and **Supplementary Table 3**). For instance, *NUMA1* encodes nuclear mitotic apparatus protein 1, involved in mitotic prometaphase of the cell cycle. Its short isoform behaved as a putative tumor suppressor through regulating the expression of MYB proto-oncogene like 2 (*MYBL2*) and played an important role in the cell cycles and cancer relapse and drug resistance (Qin et al., 2017). *ZNF143* positively regulates tumor growth through transcriptional regulation of DNA replication and cell-cycle-associated genes (such as *CDC6*, *PLK1*, and *MCM*s) in multiple solid tumors, including in LCs (Izumi et al., 2010). *MUC16*, encodes mucin 16, a cell surface–associated protein, CA125, which is a biomarker in various cancers. The overexpression of *MUC16* induced by gene mutations was reported to affect LC cells, increasing their resistance to cisplatin and gemcitabine, promoting their growth, and enhancing their migration and invasion by downregulation of p53 (Lakshmanan et al., 2017; Kanwal et al., 2018). *MOV10*, encoding Mov10 RISC complex RNA helicase, was reported to be highly expressed with *POLR2A*, *MAPK3*, and *XAB2* in 95% of 54 lung adenocarcinoma (LUAD) cases with poor prognosis (Mao et al., 2020).

For P5, although there were hundreds of mutations identified in the CTCs and primary tumor specimens, only four mutations (VAF = 0.11—0.28) were shared ones (**Figure 3F**), probably due to sampling bias in the progressive specimen. Two important genes found in the CTCs and primary and progressive tumors: *LZTR1* (a frameshift deletion at p.T7fs) of the Wnt pathway and a transcription factor gene *PELP1* of the estrogen receptor (ER) signaling pathway, which both closely relate to cancer metastasis reported in LCs (Słowikowski et al., 2015; Motta et al., 2019; Wang and Bao, 2020). Two other mutations shared by the primary tumor and CTCs were cancer-driver gene *TP53* (splicing mutation at c.97-1G > T) of the p53/CPF pathway and *POLR2A* of NER (**Supplementary Table 3**).

For P6, many consistent mutations (76.9%) between primary and progressive hydrothorax specimens were observed (**Figure 3G**). Mutations of *BPTF*, encoding a bromodomain PHD-finger transcription factor, was identified in primary tumor and CTCs. It interacts with *SMAD2* and has functions of chromatin reorganization and transcriptional regulation through the Wnt signaling pathway downstream of the TGFβ pathway (Liu et al., 2011), indicating its potential role of tumor metastasis and relapse. The *ELP3* mutation was identified in both paired tumors and CTCs. It encodes the catalytic subunit 3 of

the histone acetyltransferase elongator complex and contributes to transcript elongation and protein translation as well as chromatin organization and chromatin regulation/acetylation. Its overexpression promotes the migration and invasion of HCC and LC cells through the PI3K/AKT signaling pathway, suggesting elongator-driven metastasis in LC relapse and drug resistance (Xu et al., 2018b).

For P7, a dozen of the 53 somatic mutations shared between primary and progressive tumors (**Figure 3H**), such as *EGFR* (including hot spot mutation site p.L858R, and a non-canonical mutation p.D1014G), *TP53* and *SMAD4* mutations, indicating consistent tumor evolution in this patient. The mutations shared in CTCs and progressive tumors included *CSMD3* and *PON1*. *CSMD3* encodes CUB and Sushi multiple domains 3, a tumor suppressor. Its mutation is proved to resist to etoposide in SCLC (Qiu et al., 2019). *PON1* encodes paraoxonase 1, a member of the paraoxonase family of enzymes and exhibits lactonase and ester hydrolase activity. It is reported to increase *Nanog* expression by genome-RNAi experiment and contributes to cancer stem cell features.[3]

For P9, 137 somatic mutations were shared between primary and progressive tumors (**Figure 3J**), but only two mutations were shared with CTCs: DNA replication regulation gene *ORC1* and a transmembrane protein gene *ST6GAL2*. *ORC1* encodes origin recognition complex subunit 1, one of subunits of protein complex essential for the initiation of the DNA replication in eukaryotic cells. It is reported to interact with *CDC6* and *KAT7/HBO1* to regulate the cell cycle and be a cancer stem cell feature-related gene in LUAD (Zhang et al., 2020).

DISCUSSION

In addition to Pemetrexed, platinum-based chemotherapeutics are used for NSCLCs without targeted therapies. They mainly inhibit the division of cancer cells by causing DNA replication disorders and have a wide range of clinical applications in cancer therapies (Relling and Evans, 2015; Dilruba and Kalayda, 2016). However, the drug resistance effect is a main barrier to its treatment efficiency. CTCs are regarded as the seeds of tumor cells detached from primary tumors, and they migrate through the blood circulating system to distant locations and grow up to be metastatic tumors. The development of NGS technology help to provide large genetic information in CTCs isolated from drug-resistant patients for us to understand the drug-resistance mechanisms underlying tumor progression and explore new drug targets.

In this study, GO and KEGG enrichment analysis on the WES data showed SNVs/Indels harbored by CTCs had a distinct and relatively narrow functional profile focusing on the calcium/calmodulin binding membrane and signaling pathway as well as some essential genes of pathways in cancer in comparison with those of primary and progressive specimens (**Supplementary Table 4**). Furthermore, the SNVs/Indels shared among CTCs and primary and progressive specimens in the patients with progressive lung tumor tissue (33.3%) and hydrothorax specimens (4.4–15.6%; **Figures 3D,E,G**) were more than those with metastasis lymph nodes (0.6–11.8%; **Figures 3B,C,F,H–J**), agreeing that distant metastasis is carried by CTCs through fundamentally different evolutionary mechanisms from those of lymph node metastasis (Reiter et al., 2020).

Tumor cellular detoxification, proliferation, or apoptosis through coordination of DNA damage repair as well as drug hydration and dissociation and transmembrane and signaling transduction systems result in different responses to platinum drug treatment. Any gene mutations impairing these processes would definitely influence therapy efficiency. Genetic variations of different individuals are revealed to associate with platinum-based chemotherapy response and drug toxicity in different LC patients (Cui et al., 2017). Our studies also reveal complex and patient-specific somatic mutation features in the platinum drug–resistant advanced NSCLCs, particularly harbored by CTCs (**Figure 4**). The most frequent mutation genes were *EGFR* and *TP53*, which is previously reported in other drug-resistant cancers (Ye et al., 2016). In addition, many chemotherapy-drug transmembrane and metabolism genes, including *ABCC11*, *ABCG2*, *CYP26B1*, *CYP2C9*, *CYP2C19*, *ST6GAL2*, *RRAD*, etc., were identified in primary or progressive specimens, but only a few of them were detected in CTCs. Contrarily, the functions of shared mutations carried by CTCs were frequently enriched to cancer genes of regulation of RTK/RAS signaling; cell cycle and apoptosis; and TGFβ signaling pathways, particularly related to stem cell features, such as *SHKBP1* in P2, *PON1* in P7, and *ORC1* in P9 as well as *NUMA1*, *ZNF143* and *MUC16* of P4, *PELP1* of P5, and possibly *ELP3* and *BPTF* of P6 (**Table 2**), indicating the difference of intrinsic or acquired resistance between metastatic lymph nodes and CTCs in drug-resistant patients. The majority of these genes are derived from primary tumor samples (in 71.4% of patients except for two patients without the primary tumor samples) and a few genes (in 28.6% patients) might acquire mutations during tumor distance metastasis. The mutations or expressions of these genes are reported to have essential roles in cancer progression under platinum-based drug treatment and worthy to explore for new drug targets for antiplatinum drug resistance.

CONCLUSION

Individual dissection of mutational profiles of CTCs and paired tumor samples of platinum-based patients based on NGS technology not only demonstrate a genetic spectrum through the blood circulating system distinct from that of the lymph circulating system, but also reveal recurrent essential mutation genes related to cell proliferation and stem cell features, which facilitates our understanding of the molecular mechanisms underlying platinum-drug resistance and cancer metastases through CTCs. Therefore, it is practical to use the CTC-NGS strategy for new drug target exploration if progressive tumor specimens are unavailable in clinical practice, particularly for advanced patients.

[3]https://www.genecards.org/cgi-bin/carddisp.pl?gene=PON1

DATA AVAILABILITY STATEMENT

All the sequencing raw data used in this study have been uploaded to the Genome Sequence Archive depository database of National Genomics Data Center, China National Center for Bioinformation (https://ngdc.cncb.ac.cn/gsa) with the submission number HRA001083.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Shanghai Chest Hospital (No. KS1740). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

RL designed the research and supervised the study. YW, XL, and HL performed the research. YC, BL, RW, and SW analyzed the data. BY, AG, WW, and AH managed the samples. YC, YW, SW, and RL wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (No. 81773273).

ACKNOWLEDGMENTS

We would like to thank Xiaobo Gong and Jie Liu of Shanghai Jiao Tong University on the helps of isolation and identification of CTCs.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.722078/full#supplementary-material

REFERENCES

Abouleila, Y., Onidani, K., Ali, A., Shoji, H., Kawai, T., Lim, C., et al. (2019). Live single cell mass spectrometry reveals cancer-specific metabolic profiles of circulating tumor cells. *Cancer Sci.* 110, 697–706. doi: 10.1111/cas.13915

Alix-Panabières, C., and Pantel, K. (2013). Circulating tumor cells: liquid biopsy of cancer. *Clin. Chem.* 59, 110–118.

Baccelli, I., Schneeweiss, A., Riethdorf, S., Stenzinger, A., Schillert, A., Vogel, V., et al. (2013). Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat. Biotechnol.* 31, 539–544.

Carlisle, J. W., Steuer, C. E., Owonikoko, T. K., and Saba, N. F. (2020). An update on the immune landscape in lung and head and neck cancers. *CA Cancer J. Clin.* 70, 505–517. doi: 10.3322/caac.21630

Carter, L., Rothwell, D. G., Mesquita, B., Smowton, C., Leong, H., Fernandez-Gutierrez, F., et al. (2017). Molecular analysis of circulating tumor cells identifies distinct copy-number profiles in patients with chemosensitive and chemorefractory small-cell lung cancer. *Nat. Med.* 23, 114–119. doi: 10.1038/nm.4239

Chemi, F., Rothwell, D. G., McGranahan, N., Gulati, S., Abbosh, C., Pearce, S. P., et al. (2019). Pulmonary venous circulating tumor cell dissemination before tumor resection and disease relapse. *Nat. Med.* 25, 1534–1539. doi: 10.1038/s41591-019-0593-1

Cui, J. J., Wang, L. Y., Zhu, T., Gong, W. J., Zhou, H. H., Liu, Z. Q., et al. (2017). Gene-gene and gene-environment interactions influence platinum-based chemotherapy response and toxicity in non-small cell lung cancer patients. *Sci. Rep.* 7:5082.

Dilruba, S., and Kalayda, G. V. (2016). Platinum-based drugs: past, present and future. *Cancer Chemother. Pharmacol.* 77, 1103–1124. doi: 10.1007/s00280-016-2976-z

Drapkin, B. J., George, J., Christensen, C. L., Mino-Kenudson, M., Dries, R., Sundaresan, T., et al. (2018). Genomic and functional fidelity of small cell lung cancer patient-derived Xenografts. *Cancer Discov.* 8, 600–615.

Girotti, M. R., Gremel, G., Lee, R., Galvani, E., Rothwell, D., Viros, A., et al. (2016). Application of sequencing, liquid biopsies, and patient-derived xenografts for personalized medicine in melanoma. *Cancer Discov.* 6, 286–299. doi: 10.1158/2159-8290.cd-15-1336

Hodgkinson, C. L., Morrow, C. J., Li, Y., Metcalf, R. L., Rothwell, D. G., Trapani, F., et al. (2014). Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. *Nat. Med.* 20, 897–903.

Hou, H. W., Warkiani, M. E., Khoo, B. L., Li, Z., Soo, R., Tan, D., et al. (2013). Isolation and retrieval of circulating tumor cells using centrifugal forces. *Sci. Rep.* 3:1259.

Izumi, H., Wakasugi, T., Shimajiri, S., Tanimoto, A., Sasaguri, Y., Kashiwagi, E., et al. (2010). Role of ZNF143 in tumor growth through transcriptional regulation of DNA replication and cell-cycle-associated genes. *Cancer Sci.* 101, 2538–2545. doi: 10.1111/j.1349-7006.2010.01725.x

Jia, Q., Chiu, L., Wu, S., Bai, J., Peng, L., Zheng, L., et al. (2020). Tracking neoantigens by personalized circulating tumor DNA sequencing during checkpoint blockade immunotherapy in non-small cell lung cancer. *Adv. Sci. (Weinh)* 7:1903410. doi: 10.1002/advs.201903410

Jia, Z., Wang, M., Li, S., Li, X., Bai, X., Xu, Z., et al. (2018). U-box ubiquitin ligase PPIL2 suppresses breast cancer invasion and metastasis by altering cell morphology and promoting SNAI1 ubiquitination and degradation. *Cell Death Dis.* 9:63.

Kanwal, M., Ding, X. J., Song, X., Zhou, G. B., and Cao, Y. (2018). MUC16 overexpression induced by gene mutations promotes lung cancer cell growth and invasion. *Oncotarget* 9, 12226–12239. doi: 10.18632/oncotarget.24203

Knijnenburg, T. A., Wang, L., Zimmermann, M. T., Chambwe, N., Gao, G. F., Cherniack, A., et al. (2018). Genomic and molecular landscape of DNA damage repair deficiency across the cancer genome Atlas. *Cell Rep.* 23, 239–254.e236.

Lakshmanan, I., Salfity, S., Seshacharyulu, P., Rachagani, S., Thomas, A. S., Das, S., et al. (2017). MUC16 regulates TSPYL5 for lung cancer cell growth and chemoresistance by suppressing p53. *Clin. Cancer Res.* 23, 3906–3917. doi: 10.1158/1078-0432.ccr-16-2530

Letouzé, E., Shinde, J., Renault, V., Couchy, G., Blanc, J., Tubacher, E., et al. (2017). Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* 8:1315.

Lin, E., Cao, T., Nagrath, S., and King, M. R. (2018). Circulating tumor cells: diagnostic and therapeutic applications. *Annu. Rev. Biomed. Eng.* 20, 329–352. doi: 10.1146/annurev-bioeng-062117-120947

Liu, Z., Lin, X., Cai, Z., Zhang, Z., Han, C., Jia, S., et al. (2011). Global identification of SMAD2 target genes reveals a role for multiple co-regulatory factors in zebrafish early gastrulas. *J. Biol. Chem.* 286, 28520–28532. doi: 10.1074/jbc.m111.236307

Mao, C. G., Jiang, S. S., Shen, C., Long, T., Jin, H., Tan, Q., et al. (2020). BCAR1 promotes proliferation and cell growth in lung adenocarcinoma via upregulation of POLR2A. *Thorac. Cancer* 11, 3326–3336. doi: 10.1111/1759-7714.13676

Morrow, C. J., Trapani, F., Metcalf, R. L., Bertolini, G., Hodgkinson, C., Khandelwal, G., et al. (2016). Tumourigenic non-small-cell lung cancer mesenchymal circulating tumour cells: a clinical case study. *Ann. Oncol.* 27, 1155–1160. doi: 10.1093/annonc/mdw122

Motta, M., Fidan, M., Bellacchio, E., Pantaleoni, F., Schneider-Heieck, K., Coppola, S., et al. (2019). Dominant Noonan syndrome-causing LZTR1 mutations specifically affect the Kelch domain substrate-recognition surface and enhance RAS-MAPK signaling. *Hum. Mol. Genet.* 28, 1007–1022. doi: 10.1093/hmg/ddy412

Qin, W. S., Wu, J., Chen, Y., Cui, F. C., Zhang, F. M., Lyu, G. T., et al. (2017). The short isoform of nuclear mitotic apparatus protein 1 functions as a putative tumor suppressor. *Chin. Med. J.* 130, 1824–1830. doi: 10.4103/0366-6999.211535

Qiu, Z., Lin, A., Li, K., Lin, W., Wang, Q., Wei, T., et al. (2019). A novel mutation panel for predicting etoposide resistance in small-cell lung cancer. *Drug Design Dev. Ther.* 13, 2021–2041. doi: 10.2147/dddt.s205633

Reiter, J. G., Hung, W. T., Lee, I. H., Nagpal, S., Giunta, P., Degner, S., et al. (2020). Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases. *Nat. Genet.* 52, 692–700. doi: 10.1038/s41588-020-0633-2

Relling, M. V., and Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature* 526, 343–350.

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., et al. (2018). Oncogenic signaling pathways in the cancer genome Atlas. *Cell* 173, 321–337.e310.

Shang, R., Wang, J., Sun, W., Dai, B., Ruan, B., Zhang, Z., et al. (2016). RRAD inhibits aerobic glycolysis, invasion, and migration and is associated with poor prognosis in hepatocellular carcinoma. *Tumour Biol.* 37, 5097–5105.

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2021. *CA Cancer J. Clin.* 71, 7–33.

Słowikowski, B. K., Gałęcki, B., Dyszkiewicz, W., and Jagodziński, P. P. (2015). Increased expression of proline-, glutamic acid- and leucine-rich protein PELP1 in non-small cell lung cancer. *Biomed. Pharmacother.* 73, 97–101. doi: 10.1016/j.biopha.2015.05.015

Taylor, M. D., Nagji, A. S., Bhamidipati, C. M., Theodosakis, N., Kozower, B. D., Lau, C. L., et al. (2012). Tumor recurrence after complete resection for non-small cell lung cancer. *Ann. Thorac. Surg.* 93, 1813–1820.

Uramoto, H., and Tanaka, F. (2014). Recurrence after surgery in patients with NSCLC. *Transl. Lung Cancer Res.* 3, 242–249.

Wang, D., and Bao, B. (2020). Gallic acid impedes non-small cell lung cancer progression via suppression of EGFR-dependent CARM1-PELP1 complex. *Drug Design Dev. Ther.* 14, 1583–1592. doi: 10.2147/dddt.s228123

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. doi: 10.1093/nar/gkq603

Wang, T., Wang, D., Zhang, L., Yang, P., Wang, J., Liu, Q., et al. (2019). The TGFβ-miR-499a-SHKBP1 pathway induces resistance to EGFR inhibitors in osteosarcoma cancer stem cell-like cells. *J. Exp. Clin. Cancer Res. CR* 38:226.

Xu, Y., Zhang, F., Pan, X., Wang, G., Zhu, L., Zhang, J., et al. (2018a). Xenograft tumors derived from malignant pleural effusion of the patients with non-small-cell lung cancer as models to explore drug resistance. *Cancer Commun. (Lond)* 38:19. doi: 10.1186/s40880-018-0284-1

Xu, Y., Zhou, W., Ji, Y., Shen, J., Zhu, X., Yu, H., et al. (2018b). Elongator promotes the migration and invasion of hepatocellular carcinoma cell by the phosphorylation of AKT. *Int. J. Biol. Sci.* 14, 518–530. doi: 10.7150/ijbs.23511

Ye, S., Shen, J., Choy, E., Yang, C., Mankin, H., Hornicek, F., et al. (2016). p53 overexpression increases chemosensitivity in multidrug-resistant osteosarcoma cell lines. *Cancer Chemother. Pharmacol.* 77, 349–356. doi: 10.1007/s00280-015-2944-z

Yu, M., Bardia, A., Aceto, N., Bersani, F., Madden, M. W., Donaldson, M. C., et al. (2014). Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science (New York, NY).* 345, 216–220. doi: 10.1126/science.1253533

Zhang, Y., Tseng, J. T., Lien, I. C., Li, F., Wu, W., and Li, H. (2020). mRNAsi index: machine learning in mining lung adenocarcinoma stem cell biomarkers. *Genes (Basel)* 11:257. doi: 10.3390/genes11030257

Conflict of Interest: YW, BL, XL, RW, HL, and SW are employees of Berry Oncology Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 30 September 2021
doi: 10.3389/fgene.2021.722601

Prognostic Value of Immune-Related Multi-lncRNA Signatures Associated With Tumor Microenvironment in Esophageal Cancer

Jingjing Pang[1†], He Pan[1†], Chunxiu Yang[1], Pei Meng[1], Wen Xie[1], Jiahao Li[1], Yueying Li[1] and Shu-Yuan Xiao[1,2]**

[1]Department of Pathology, Wuhan University Center for Pathology and Molecular Diagnostics, Zhongnan Hospital of Wuhan University, Wuhan, China, [2]Department of Pathology, University of Chicago Medicine, Chicago, IL, United States

OPEN ACCESS

Edited by:
Yuriy L. Orlov,
I. M. Sechenov First Moscow State Medical University, Russia

Reviewed by:
Divya Mundackal Sivaraman,
Sree Chitra Tirunal Institute for Medical Sciences and Technology (SCTIMST), India
Anatoliy Ivashchenko,
Al-Farabi Kazakh National University, Kazakhstan
Evgenii Chekalin,
Michigan State University, United States

***Correspondence:**
Shu-Yuan Xiao
shu-yuan.xiao@uchospitals.edu,
Yueying Li
liyueying0525@hotmail.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 09 June 2021
Accepted: 03 September 2021
Published: 30 September 2021

Citation:
Pang J, Pan H, Yang C, Meng P, Xie W, Li J, Li Y and Xiao S-Y (2021) Prognostic Value of Immune-Related Multi-lncRNA Signatures Associated With Tumor Microenvironment in Esophageal Cancer. Front. Genet. 12:722601. doi: 10.3389/fgene.2021.722601

Esophageal cancer is the eighth most common cancer and the sixth leading cause of cancer death worldwide. Hence, for a better understanding of tumor microenvironment and to seek for novel molecular targets for esophageal cancer, we performed related studies on two histopathological subtypes of esophageal cancer: esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC). Bioinformatic analyses were conducted based on the RNA-seq, genomic mutation, and clinical data from TCGA and GEO cohorts. We clustered patients into high-immunity and low-immunity groups through the ssGSEA results. The ESTIMATE algorithm was used to evaluate the tumor microenvironment. Patients with high immunity in both ESCC and EAC had lower tumor purity and poor survival. Subsequently, CIBERSORT was performed to learn about the detailed difference of tumor-infiltrating lymphocytes (TILs) between high- and low-immunity patients. Specific increase of M2 macrophages and decrease of activated dendric cells can be observed in ESCC and EAC, respectively. The most enriched functions and pathways of high-immunity patients were immunoglobulin complex, MHC class II protein complex, and allograft rejection according to the GO terms and KEGG. Two prognostic immune-related multi-lncRNA risk models were constructed and validated by ROC curve and PCA in ESCC and EAC. High-risk patients in both subtypes had poor survival, advanced clinical characteristics, and higher drug susceptibility except cisplatin and sorafenib. In addition, the tumor mutation burden (TMB) was positively correlated with the risk level in the ESCC and EAC and showed distinct differences between the two subtypes. In summary, we comprehensively analyzed the tumor microenvironment for two subtypes of esophageal cancer, identified two multi-lncRNA signatures predictive for the prognosis, and explored the possibility of the signatures to forecast drug susceptibility as well as TMB for the first time. The findings may serve as a conceptual basis for innovative strategy of individualized immunotherapy for esophageal cancer.

Keywords: esophageal cancer, tumor microenvironment, long noncoding RNAs, risk score, prognosis, tumor mutation burden, drug susceptibility

INTRODUCTION

Esophageal cancer is a highly invasive malignancy with poor prognosis. According to World Health Statistics in 2018, the incidence of esophageal cancer ranked eighth and mortality ranked sixth (Gong et al., 2019). The 5-years overall survival rate is approximately 15% (Talukdar et al., 2018). Poor outcomes in patients with esophageal cancer are related to diagnosis at advanced stages and the propensity for metastases (Thrift, 2021). Esophageal cancer is generally classified into two histopathological subtypes: esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC). ESCC usually originates from esophageal squamous epithelial cells, always driven by the exposure of tobacco, alcohol, hot drink and malnutrition. In contrast, EAC develops from columnar metaplasia of the lower esophagus and is related to obesity and gastric acid reflux (Fatehi Hassanabad et al., 2020). These two subtypes have striking differences in geographical distribution possibly on account of the differences in exposure to risk factor and lifestyle. ESCC comprises the vast majority of esophageal cancer in southeastern and central Asia, southeastern Africa, and south America, whereas EAC is the predominant subtype in Northern Europe, Western Europe, North America, and Oceania, constituting approximately 46% of the global EAC with more pronounced differences in gender than ESCC (Dong et al., 2018).

Currently, the major therapeutic approaches of esophageal cancer are surgical resection and neoadjuvant chemoradiotherapy (CRT). The most commonly used biological and targeted agents in esophageal cancer included angiogenesis inhibitor ramucirumab and the inhibitors of epidermal-growth-factor receptors trastuzumab (Fatehi Hassanabad et al., 2020). However, due to the heterogeneity of esophageal cancer leading to inherent resistance to chemotherapy, and limited clinical benefits of intervention or targeted therapy, the survival and prognosis of advanced patients remain disappointing. Nowadays, immunotherapy has become a promising treatment approach, which aims to activate the immune system and rely on its intrinsic immune function to kill tumor cells. The immunotherapy includes chimeric antigen receptor T cells (CAR-T) therapy, immune-checkpoint blockade (ICB), oncolytic virus, and tumor vaccines (Baba et al., 2020). ICB has shown strong anti-tumor activity in solid tumors such as malignant melanoma, non-small cell lung cancer, renal clear cell carcinoma, and prostate cancer (Riley et al., 2019). Anti-PD-1/PD-L1 antibodies has brought a historic revolution for immunotherapy. Pembrolizumab has been approved by the U.S. Food and Drug Administration (FDA) to treat PD-L1 positive patients who have progressive disease after second-line therapies. Unfortunately, this anti-PD-1 antibody failed to improve the treatment efficacy in patients with advanced PD-L1-positive esophageal cancer (Shitara et al., 2018). With deeper research on the immunotherapy of esophageal cancer, more and more evidence showed that the complicated tumor microenvironment of esophageal cancer contributed to the intervention of anti-tumor immunoregulation or immunotherapy of esophageal cancer. Nevertheless, the exact mechanisms were not yet elucidated. Thus, a better understanding of antitumor immunity and tumor microenvironment is of utmost importance to improve the efficiency of immunotherapy.

Tumor microenvironment refers to the cellular environment in which the tumor develops, comprising tumor cells, endothelial cells, fibroblasts, immune cells, cytokines, growth factors, and extracellular matrix (Wu and Dai, 2017). Tumor cells can functionally secrete various cytokines, chemokines, and other factors to sculpt the microenvironment resulting in the alteration of the surrounding cells to intricately influence the occurrence and development of tumor. Tumor-associated macrophages, cytokines, IL-1, and complement have emerged as promoters in tumorigenesis, while myeloid cells and innate lymphoid cells are recognized as the potential tumor suppressor. Growing evidence suggested that the tumor microenvironment plays a pivotal role in regulating immune responses, facilitating immune escape, promoting angiogenesis, and inducing metastasis, contributing to a far-reaching impact on the effectiveness of immunotherapy (Meurette and Mehlen, 2018; Hinshaw and Shevde, 2019; Jarosz-Biej et al., 2019). Meanwhile, tumor mutation burden (TMB) was discovered as a novel biomarker to predict the efficacy of ICB. Higher TMB was generally related to better overall survival after ICB therapy for a variety of cancers, including non-small cell lung cancer, colorectal cancer, bladder cancer, and melanoma (Bader et al., 2020). Concerning the conventional reliable markers in immunotherapy of esophageal cancer, PD-L1 immunohistochemical evaluation results revealed that it hardly accurately predicted the therapeutic response to anti-PD-1 antibody in esophageal cancer patients (Yang et al., 2020a). Based on it, TMB is emerging as an immune-response biomarker for esophageal cancer.

Recently, researches have indicated that lncRNA may also be involved in tumor microenvironment remodeling in esophageal cancer, suggesting that it is of great value to study lncRNA associated with immunity (Robinson et al., 2020). Long non-coding RNAs (lncRNAs) are single-stranded RNAs longer than 200 nucleotides without protein coding potential, which mainly functions as the regulators of chromatin dynamics and gene regulation, closely associated with transcription, translation, and epigenetic modification (Qian et al., 2019). Aberrant expression, mutations, and SNPs of lncRNA are supposed to be correlated to tumorigenesis and metastasis (Wang et al., 2021). Genome-wide association studies have identified a large number of lncRNAs that may serve as biomarkers and therapeutic targets for esophageal cancer.

In this study, we aimed to plot the comprehensive landscape of tumor microenvironment and explore the prognostic immune-related multi-lncRNA signatures in ESCC and EAC. We hope that our findings will help elucidate the pathological mechanism partly and make further contribution to esophageal cancer.

MATERIALS AND METHODS

Data Sources

The publicly available esophageal cancer patient datasets were directly downloaded from The Cancer Genome Atlas (TCGA) data portal (https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.

htm) and Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/), which contained RNASeqV2 normalized gene expression data of 135 ESCC samples (TCGA: 95, GEO: 53) and 143 EAC samples (TCGA: 87, GEO: 56) in total. The datasets of GEO originated from GSE54994 and GSE20154. The tumor somatic mutation data of two subtypes of esophageal cancer were also obtained from TCGA and GEO. Clinical information for 185 TCGA esophageal cancer cohorts was downloaded from UCSC Xena (http://xena.ucsc.deu/).

Single Sample Gene Set Enrichment Analysis

We calculated the enrichment levels of 29 immune-associated datasets in each esophageal cancer sample in the form of ssGSEA scores. These 29 immune signatures containing diverse immune cell types, function, and pathways were obtained from previous publications (Kobayashi et al., 2020) (**Supplementary Table S1**). Unsupervised hierarchical clustering was performed to classify patients into two subtypes: high immunity and low immunity."GSVA" R package was used to do the cluster analysis.

Analysis for Tumor Environment and Immune Infiltration

Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) is a tool for predicting tumor purity and presence of infiltrating stromal/immune cells in tumor samples. ESTIMATE algorithm was executed on the basis of ssGSEA results and generated three scores: stromal score (that captures the presence of stroma in tumor tissue), immune score (that represents the infiltration of immune cells in tumor tissue), and estimate score (that infers tumor purity) by "ESTIMATE" R package (Kang et al., 2020).

CIBERSORT algorithm is an R/Web-based tool for deconvolving the expression matrix of human immune cells based on linear support vector regression. Gene expression profiles of 22 common immune cells were downloaded as reference marker from CIBERSORT (https://cibersortx.stanford.edu/). The abundances of the 22 immune cells in esophageal cancer patients were calculated with "CIBERSORT" R package (Yang et al., 2021).

The immune infiltration analysis between high- and low-risk patients was conducted based on the file named "infiltration_estimation_for_tcga.csv" downloaded from TCGA including the immune infiltration data calculated by TIMER, CIBERSORT, CIBERSORT-ABS, EPIC, quanTIseq, MCP-counter, and xCell.

Building Risk Prediction Model for Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma

Univariate Cox proportional risk regression analysis was performed for each immune-related lncRNA with survival data. Least Absolute Shrinkage and Selection Operator (LASSO)-penalized Cox regression was utilized one step forward to obtain the best candidates of multi-lncRNA for predicting prognosis through the use of the "glmnet" package in R software. Afterwards, a risk score model of the prognostic multi-lncRNA was established according to the following formula: Lasso Risk core = $\sum_{i=1}^{n} Coef_i * Exp_i$.

Verification Study

In the validation of the risk model, receiver operating characteristic (ROC) curve and Principal Components Analysis (PCA) were conducted. We used the R package "survival ROC" for time-dependent ROC curve analysis. PCA served as a dimensionality reduction algorithm, utilizing the matrix of normalized gene counts of lncRNA in the risk model. Through orthogonal transformation, we maximized accuracy and minimized the error of overfitting, establishing the correlation between patients with high and low risk. The outcome was visualized using "scatterplot3d" package in R software.

Evaluation of Drug Sensitivity

IC_{50} represented the concentration necessary for 50% inhibition. We calculated IC_{50} of drugs through "pRRophetic" R package and its dependencies including "car, ridge preprocessCore, genefilter and sva", which contained the effect information of 138 drugs. The boxplot was plotted by the use of "ggplot2" R package (Song et al., 2020).

Calculation of Tumor Mutation Burden Scores

Estimation was practiced to count the average number of somatic mutations in tumor genome including coding base substitutions insertions or indels per megabase (Mb) of the sequence examined based on the annotated list from TCGA-ESCA and GEO. We took 38 Mb as a routine value of the length of the human exon and divided the total mutation counts of missense, nonstop, nonsense, and frameshift number by 38 to compute TMB scores. "maftools" R package was employed to draw the waterfall plotting illustrating the relationship between risk scores and TMB in esophageal cancer patients (Kang et al., 2020).

Statistical Analysis

Statistical analysis was performed with R software version 4.0.4 on R Studio and GraphPad Prism eight software. The RNASeqV2 normalized gene expression and somatic mutation data from TCGA and GEO was merged by the "limma" R package and batch effect was removed by the "sva" R package. The differentially expressed immune-related lncRNAs were screened out with the application of the "limma" R package. Cox regression and survival analysis were conducted through "survival" and "survminer" R package. Hazard ratios for univariate and multivariate analysis were calculated by Cox proportional hazards regression model. The "pheatmap" R package was used for plotting heatmaps in analyses of cluster and risk score model. Wilcoxon rank-sum test was used to compare the difference of two groups of quantitative data; $p < 0.05$ was considered significant. Fisher's exact test was used to

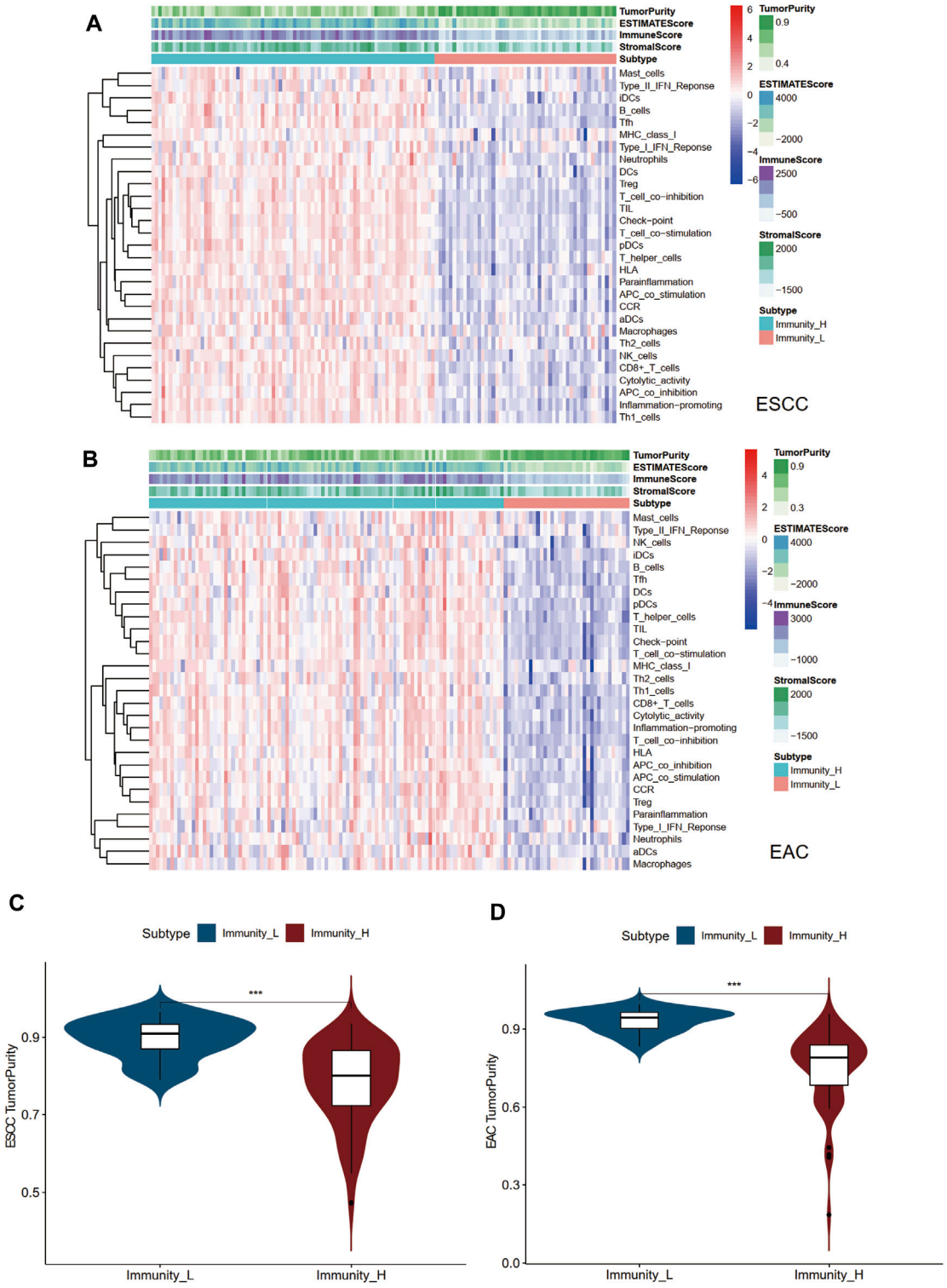


FIGURE 1 | Landscape of tumor immune microenvironment in ESCC and EAC. **(A)** Heatmap of the tumor purity, ESTIMATE scores, stromal scores, immune scores, and tumor-infiltrating lymphocytes in Immunity_L and Immunity_H groups in ESCC. **(B)** Heatmap of the tumor purity, ESTIMATE scores, stromal scores, immune scores, and tumor-infiltrating lymphocytes in Immunity_L and Immunity_H groups in EAC. **(C)** Difference of tumor purity between Immunity_L and Immunity_H groups in ESCC. **(D)** Difference of tumor purity between Immunity_L and Immunity_H groups in EAC. ESCC, esophageal squamous cell carcinoma; EAC, esophageal adenocarcinoma; ssGSEA, single sample Gene Set Enrichment Analysis; ESTIMATE, Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

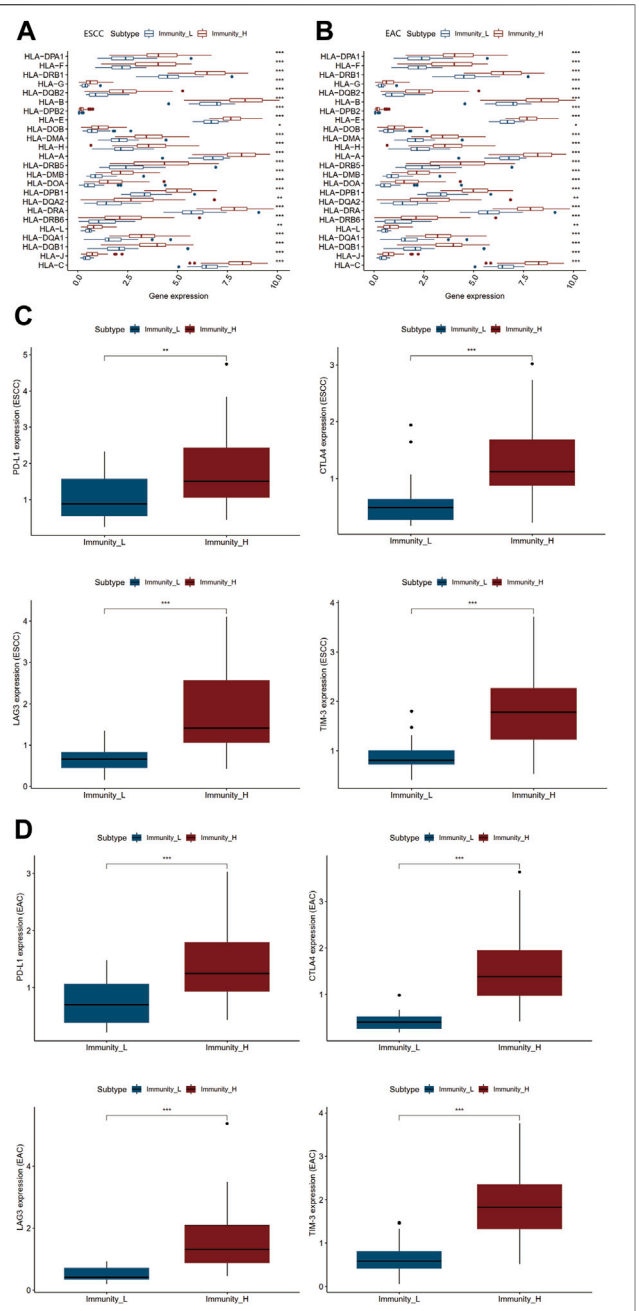


FIGURE 2 | Differential genetic analyses of two immunity clusters in ESCC and EAC. **(A)** Differential expression of HLA family genes in Immunity_L and Immunity_H groups of ESCC. **(B)** Differential expression of HLA family genes in Immunity_L and Immunity_H groups of EAC. **(C)** Differential expression of PD-L1, CTLA4, LAG3, and TIM-3 in Immunity_L and Immunity_H groups of ESCC. **(D)** Differential expression of PD-L1, CTLA4, LAG3, and TIM-3 in Immunity_L and Immunity_H groups of EAC. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

analyze the ratio difference of immunity and risk cluster. The Gene ontology (GO) terms and KEGG pathways were analyzed by the R package "clusterProfiler", which were identified by a threshold of $p < 0.05$.

RESULTS

Tumor Immune Microenvironment in Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma

Two immune clusters of esophageal cancer, namely, Immunity high (Immunity_H) and Immunity low (Immunity_L), were grouped according to the enrichment scores of infiltrating immune cells, immunity functions, and pathways between ESCC and EAC samples by the ssGSEA (**Supplementary Table S2**). Then, following the ESTIMATE algorithm, tumor purity, immune score, stromal score, and ESTIMATE score were calculated to further explore the tumor microenvironment. As a result, the stromal, immune, and ESTIMATE scores of the Immunity_H group were markedly higher than those in the Immunity_L group in both ESCC and EAC (**Figures 1A,B**). Conversely, the Immunity_L group scored higher in tumor purity (**Figures 1C,D**). The heatmap demonstrated the differential stromal/immune cell infiltration between the two clusters respectively in ESCC and EAC (**Supplementary Table S3**).

Expression Level of Human Leukocyte Antigens Family Genes and Immune Checkpoints

Human leukocyte antigens (HLAs) mostly encoded by human major histocompatibility complex (MHC) were deeply involved in the resistance to foreign pathogens and immunological responses in a variety of pathological processes. Immune checkpoints have been proved to be crucial therapeutic targets for ICB (Wu et al., 2020a). We inspected the differential HLA genes and identified four immune checkpoint genes PD-L1, CTLA4, LAG3, and TIM-3 strongly associated with ESCC and EAC (Huang and Fu, 2019). As shown in **Figures 2A–D**, the expression levels of both HLA-related genes and immune checkpoints were notably elevated in the Immunity_H group in ESCC and EAC (**Supplementary Table S4**).

Prognostic Features of the Two Immune Clusters in Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma

The survival data of patients with ESCC and EAC were used for overall survival analysis. The survival curve was plotted in **Figures 3A,B** showing that Immunity_L patients had longer survival than Immunity_H patients. The same situations occurred in ESCC and EAC (**Supplementary Table S5**).

Tumor-Infiltrating Lymphocytes in the Two Immune Clusters in Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma

In order to investigate the differences of the tumor-infiltrating lymphocytes (TILs) between the two immune clusters,

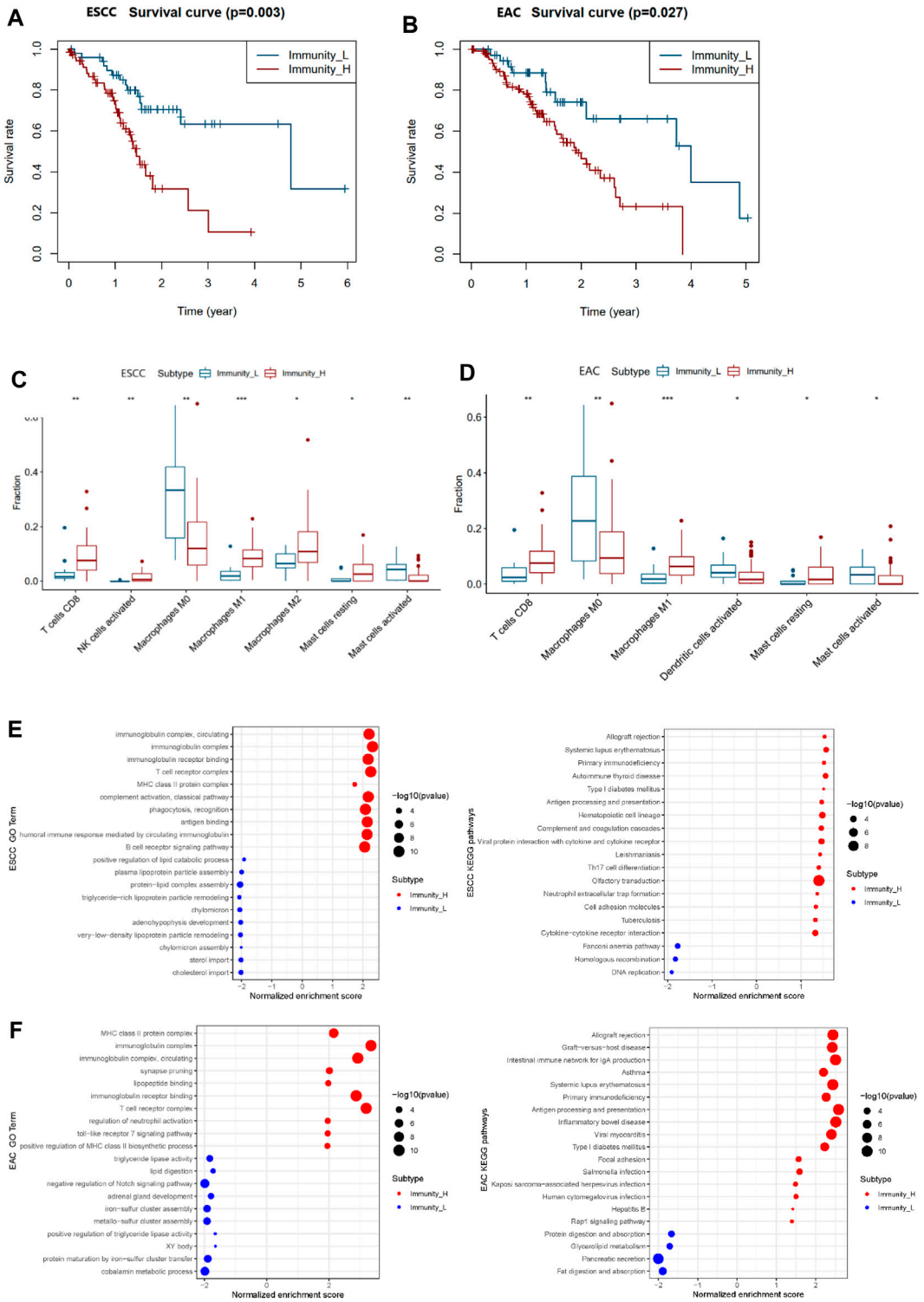
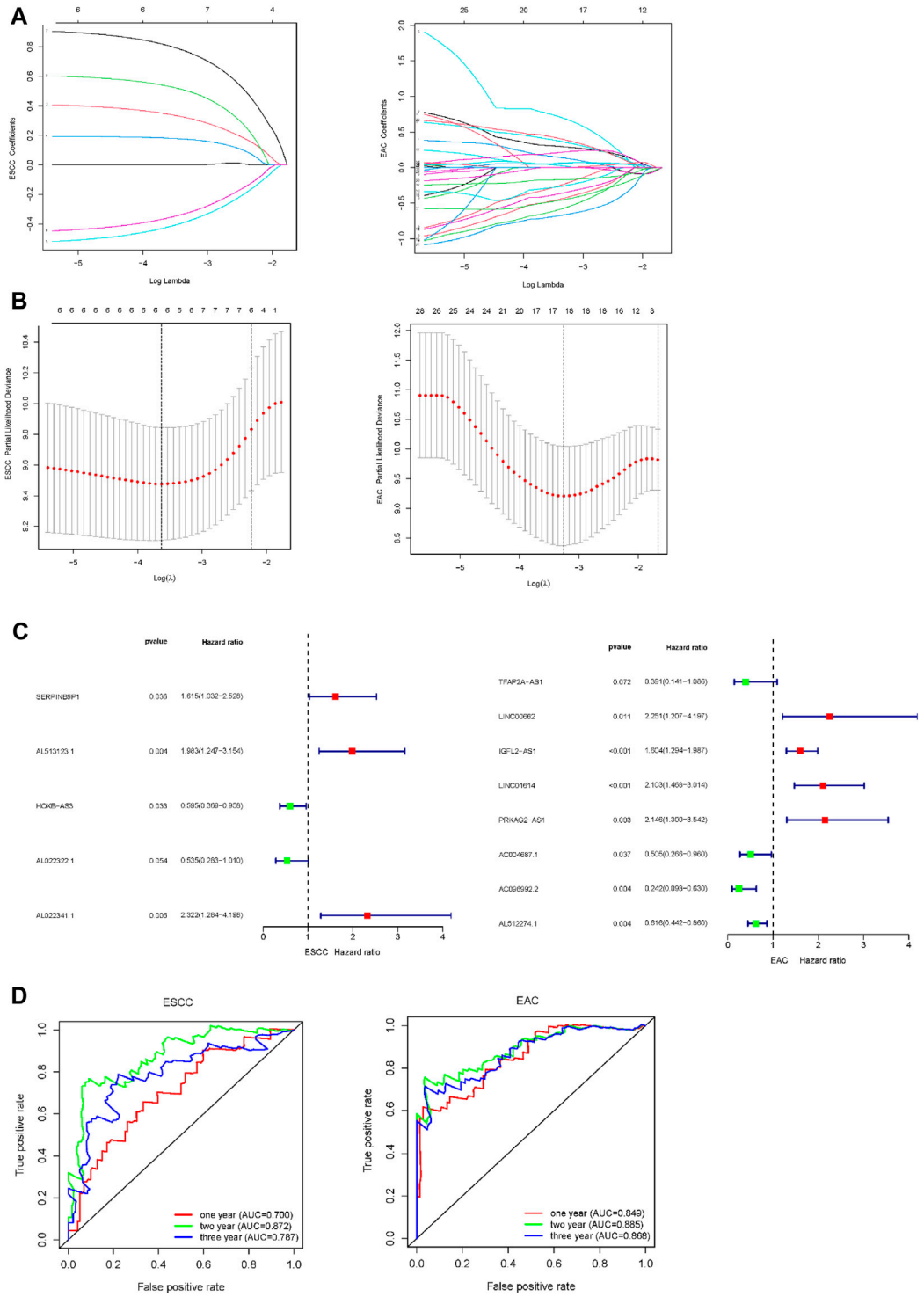


FIGURE 3 | Differential survival, TILs, and enriched functions and pathways of two immunity clusters in ESCC and EAC. **(A)** Survival curve of Immunity_L and Immunity_H groups in ESCC. **(B)** Survival curve of Immunity_L and Immunity_H groups in EAC. **(C)** Boxplot of differential TILs in Immunity_H and Immunity_L in ESCC. **(D)** Boxplot of differential TILs in Immunity_H and Immunity_L in EAC. **(E)** The most enriched GO terms and KEGG pathways between Immunity_L and Immunity_H groups in ESCC. **(F)** The most enriched GO terms and KEGG pathways between Immunity_L and Immunity_H groups in EAC. TILs, tumor-infiltrating lymphocytes; GO, Gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.



12

FIGURE 4 | Construction of the prognostic immune-related multi-lncRNA risk model in ESCC and EAC. **(A)** Lasso regression coefficient profiles of the prognostic lncRNA in ESCC and EAC. **(B)** Optimal and minimum criterion of Lasso penalization parameter (λ) used 10-fold cross-validation in ESCC and EAC cohorts. **(C)** Multi-lncRNA prognostic signatures in ESCC and EAC. **(D)** ROC curve of 1-, 2-, and 3-years overall survival for validating the capacity of predicting prognosis of the risk models. **(E)** Kaplan–Meier survival analyses in "High-risk" and "Low-risk" groups of ESCC and EAC. **(F)** Expression level of the lncRNA and patients' survival status in risk models of ESCC and EAC. **(G)** Principal component analysis (PCA) in patients of "High-risk" and "Low-risk" groups based on the immune-related lncRNA of risk models in ESCC and EAC. ROC, receiver operating characteristic. AUC, area under curve.

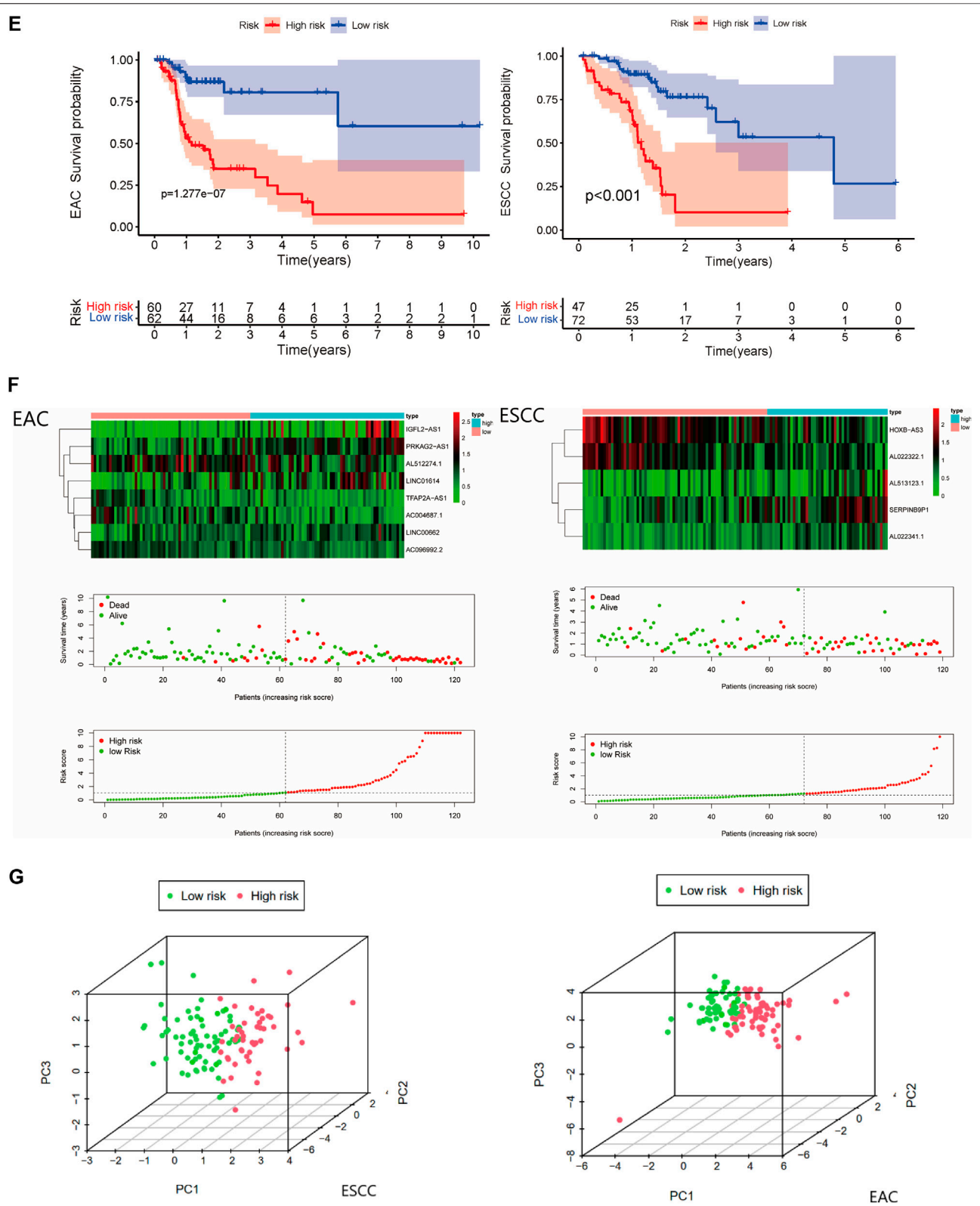

E
Risk High risk Low risk
EAC Survival probability
p=1.277e−07
Time(years)
High risk 60 27 11 7 4 1 1 1 1 1 0
Low risk 62 44 16 8 6 6 3 2 2 2 1
0 1 2 3 4 5 6 7 8 9 10
ESCC Survival probability
p<0.001
High risk 47 25 1 1 0 0 0
Low risk 72 53 17 7 3 1 0
0 1 2 3 4 5 6
F
EAC
IGFL2−AS1
PRKAG2−AS1
AL512274.1
LINC01614
TFAP2A−AS1
AC004687.1
LINC00662
AC096992.2
ESCC
HOXB−AS3
AL022322.1
AL513123.1
SERPINB9P1
AL022341.1
Dead
Alive
Survival time (years)
Patients (increasing risk socre)
High risk
low Risk
Risk score
G
Low risk High risk
PC1
PC2
PC3
ESCC
EAC

FIGURE 4 |

CIBERSORT was used to calculate the fractions of 22 TILs, respectively (**Supplementary Table S6**). Finally, we identified seven TILs of significant differences in ESCC and six TILs in EAC (**Figures 3C,D**). In ESCC, CD8+T cells, activated NK cells, M1 macrophages, M2 macrophages, and resting mast cells appeared more in the Immunity_H group than in the Immunity_L group, while the M0 macrophages and activated mast cells markedly decreased. In the Immunity_H group of EAC, CD8+T cells, M1 macrophages, and resting mast cells increased with the amounts of M0 macrophages, activated dendritic cells, and activated mast cells declining.

GO Terms and KEGG Pathways of the Two Immune Clusters in Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma

To further evaluate the similarities and differences of immunological functions between the two immune clusters of esophageal cancer, GO terms and KEGG pathway enrichment analysis was carried out. As displayed in **Figures 3E,F**, the main functions of the Immunity_H group enriched were similar in ESCC and EAC patients, including the "MHC class II protein complex", "immunoglobulin complex", "immunoglobulin receptor binding", and "T cell receptor complex". The major enriched KEGG pathways in the Immunity_H group of ESCC and EAC were associated with immune response. The most significant pathway in ESCC and EAC was "Allograft rejection". Moreover, there were some cancer-related pathways among the ESCC and EAC KEGG list, such as "cytokine-cytokine receptor interaction", "DNA replication", "focal adhesion", and "Rap1 signaling pathway" (**Supplementary Table S7**).

Construction of Multi-lncRNA Risk Assessment Model for Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma

Based on the RNA-sequencing data of TCGA and GEO samples from TCGA and GEO, we removed the genes encoding proteins and collected lncRNA expression data separately. Then, we choose the significantly differential and highly correlated with immune genes filtered by the criteria of Log|FC| > 1 and correlation coefficient $R \geq 0.4$. $p < 0.05$ was set as the significance threshold. Subsequently, we incorporated the survival data into the immune-related lncRNA expression matrix and applied univariate Cox regression analysis to characterize lncRNA with good predictive performance for prognosis. As a result, differentially expressed immune-related lncRNA associated with prognosis were screened out with $p < 0.05$ in both ESCC and EAC (**Supplementary Table S8**).

To further assess the prognostic value of immune-related lncRNA, the Lasso Cox regression analysis was conducted. We identified the variation of regression coefficients for the prognostic lncRNA and selected the optimal and minimum criterion of penalization parameter (λ) used 10-fold cross-validation (**Figures 4A,B**, **Supplementary Table S9**). According to the Lasso risk score formula, we picked out the most correlated lncRNA with survival to establish the multi-lncRNA survival risk score models (**Figure 4C**). A 5-lncRNA signature of ESCC and an 8-lncRNA signature of EAC were obtained. SERPINB9P1, AL513123.1, and AL022341.1 were the risk factors of ESCC, while HOXB-AS3 and AL022322.1 were the protective factors. In EAC, LINC00662, IGFL2-AS1, LINC01614, and PRKAG2-AS1 were the risk factors, when TFAP2A-AS1, AC004687.1, AC096992.2, and AL512274.1 were protective factors. ROC analysis of the risk assessment model was performed to estimate the specificity and time-dependent sensitivity for survival risk groups (**Figure 4D**). The 1-year, 2-years, and 3-years area under curve (AUC) for the multi-lncRNA risk score model in ESCC and EAC suggested that these two models can effectively evaluate the prognosis of esophageal cancer patients. ESCC and EAC patients were grouped into "High risk" and "Low risk" according to the cutoff values from ROC analysis. Kaplan–Meier survival curves were plotted to help visualize the survival performance more intuitively. The results showed that the overall survival of ESCC and EAC patients in the "High-risk" group was significantly poorer than that in the "Low-risk" group (**Figure 4E**). In **Figure 4F**, it presented the expression level of lncRNA and patients' survival status in the risk model. With the elevation of risk scores, the survival time declined and the number of deaths increased.

Furthermore, PCA was used to validate the accuracy of the multi-lncRNA risk models. According to **Figure 4G**, "High-risk" and "Low-risk" patients were separated completely in the multi-lncRNA risk models, which meant the risk models have good differentiability for patients.

Clinical Traits of Multi-lncRNA Signatures in Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma

To further clarify if the constituent ratio of high- and low-risk patients with esophageal cancer was proportioned in Immunity_H and Immunity_L clusters, we plotted the column charts and performed Fisher's exact test (**Figures 5A,B**). There was no significant correlation between immunity and risk level in both ESCC and EAC ($p > 0.05$) (**Supplementary Table S10**). In addition, we also explored the relationships between risk scores and clinical traits. As shown in **Figures 5C,D**, stage and alcohol were the most correlative clinical traits with the risk score in patients with ESCC, while in patients with EAC, only T (depth of tumor invasion) stages were observed relative to the risk score.

TILs of Multi-lncRNA Signatures in Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma

Depending on variate software analyzing the TILs, we identified the association between the risk scores and TILs (**Figures 5E,F**). We found that most TILs showed positive correlation with risk in

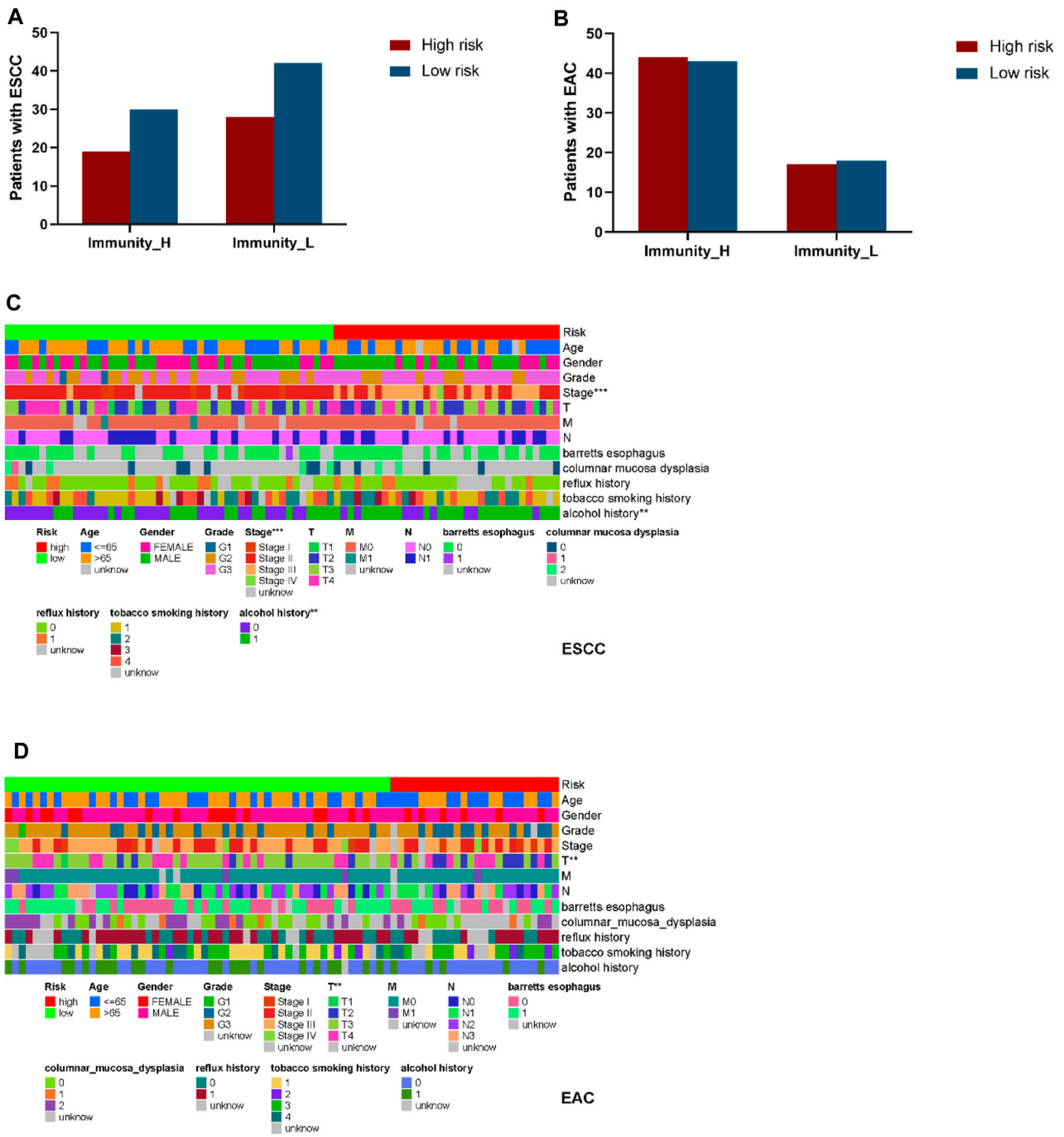


FIGURE 5 | Clinical traits and TILs correlated analyses for esophageal cancer patients. **(A)** The constituent proportion of high- and low-risk patients in immunity clusters in ESCC. **(B)** Constituent proportion of high- and low-risk patients in immunity clusters in EAC. **(C)** Heatmap of clinical traits correlated with risk in ESCC. **(D)** Heatmap of clinical traits correlated with risk in EAC. **(E)** Correlation between risk and TILs in ESCC. **(F)** Correlation between risk and TILs in EAC. T, Depth of tumor invasion; M, Metastasis; N, Lymph Node; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

both ESCC and EAC. Only common lymphoid progenitors, γδT cells, resting $CD4^{+}$ memory T cells, resting NK cells, and M0 macrophages presented negative correlation in ESCC, while the mast cells, naïve $CD4^{+}$ T cells, and M0 macrophages were negatively related to risk in EAC (**Supplementary Table S11**).

Drug Sensitivity of Multi-lncRNA Signature in Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma

To investigate the possible application of the multi-lncRNA signature to personalized treatment of esophageal cancer

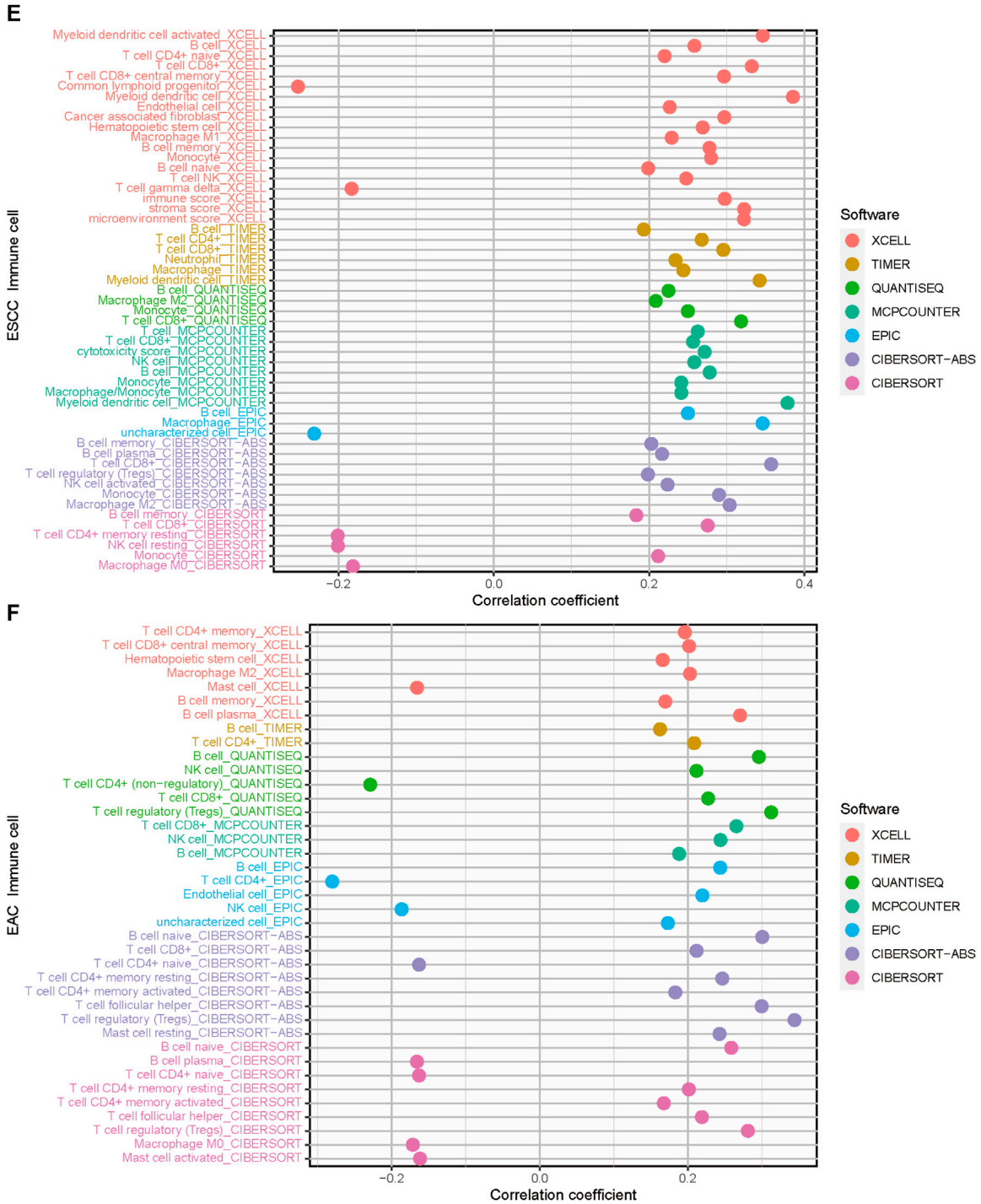


FIGURE 5 |

patients, we examined the relationship between risk scores and IC_{50} of drugs universally used or studied in the treatment of ESCC and EAC. These agents included cisplatin, sunitinib, erlotinib, gefitinib, lapatinib, and sorafenib. As shown in **Figures 6A,B**, high-risk ESCC patients appeared to be more susceptible to most of the drugs than low-risk patients. However, high-risk patients with EAC may not benefit more from cisplatin and sorafenib treatment.

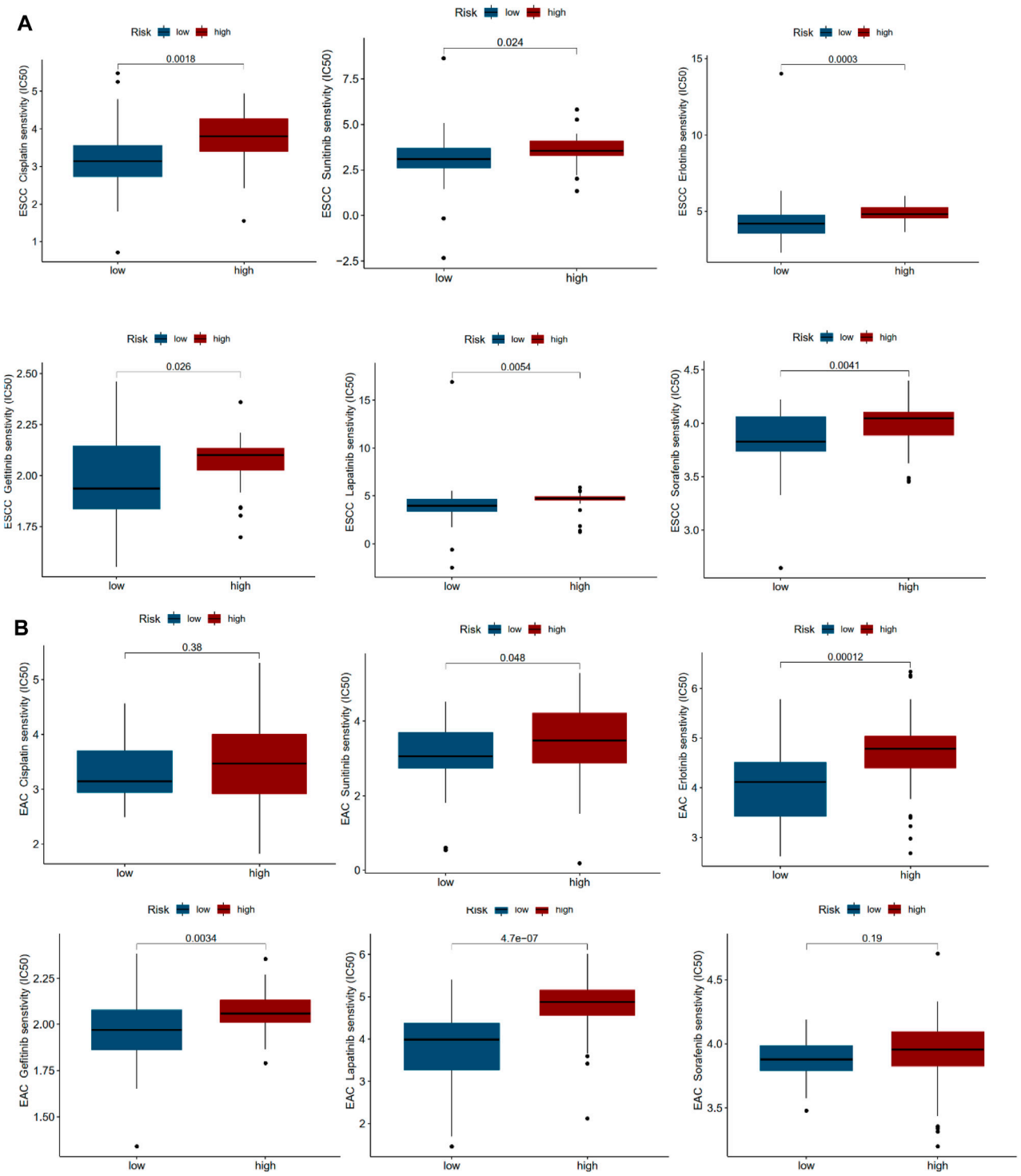


FIGURE 6 | Drug sensitivity correlated with high- and low-risk patients in ESCC and EAC. **(A)** IC_{50} values of various drugs in high- and low-risk patients with ESCC. **(B)** IC_{50} values of various drugs in high- and low-risk patients with EAC.

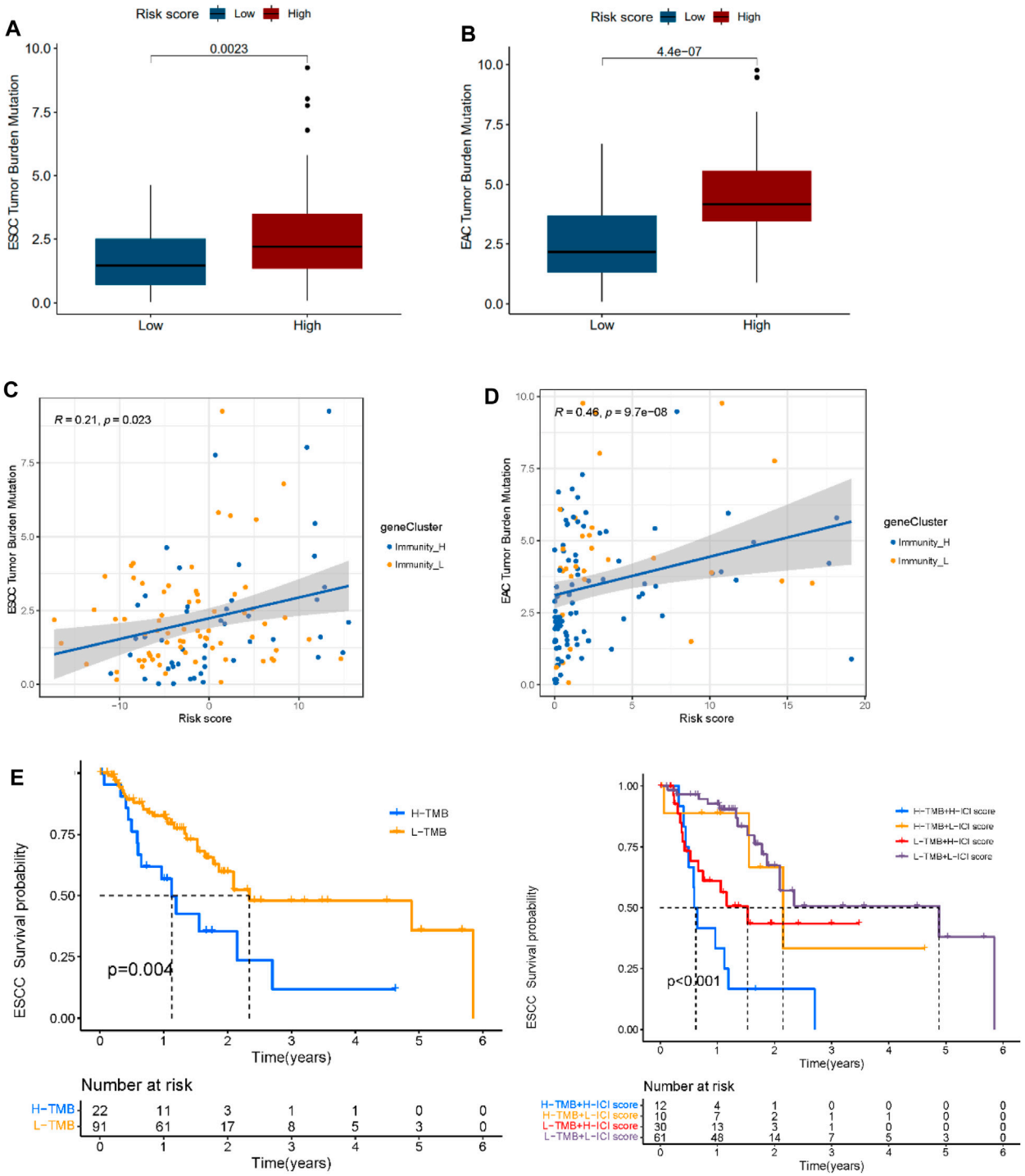


FIGURE 7 | Correlation between tumor mutation burden (TMB) and risk score. **(A)** Differential TMB in high- and low-risk groups in ESCC. **(B)** Differential TMB in high- and low-risk groups in EAC. **(C)** Linear correlation between TMB and risk score in ESCC. **(D)** Linear correlation between TMB and risk score in EAC. **(E)** Survival curves of high and low TMB groups and TMB-risk combined survival curves in ESCC. **(F)** Survival curves of high and low TMB groups and TMB-risk combined survival curves in EAC. **(G)** Waterfall plot of top 20 mutant genes in the high- and low-risk group in ESCC. **(H)** Waterfall plot of the top 20 mutant genes in the high- and low-risk group in EAC.

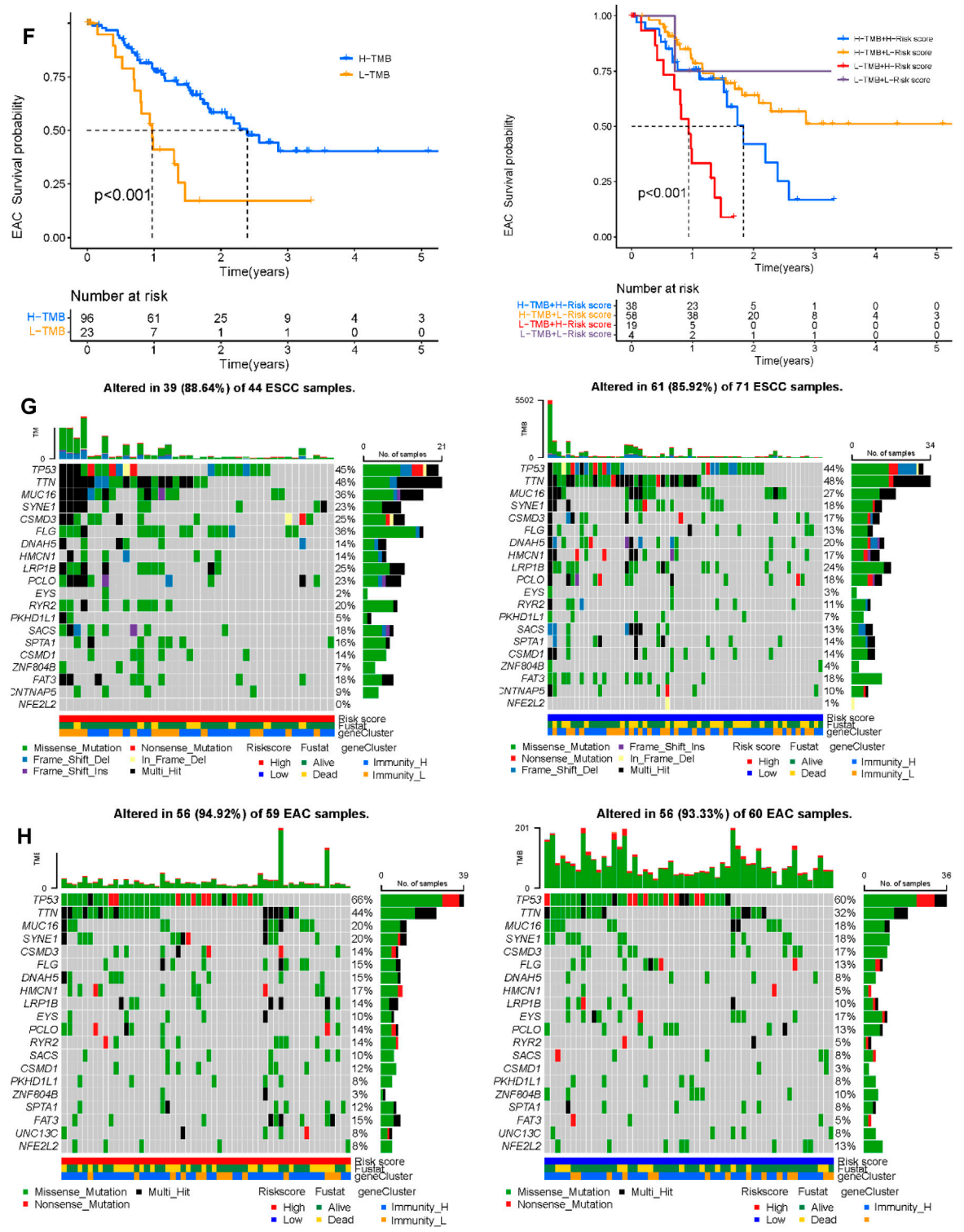

F
H-TMB
L-TMB
EAC Survival probability
p<0.001
Time(years)
Number at risk
H-TMB 96 61 25 9 4 3
L-TMB 23 7 1 1 0 0
H-TMB+H-Risk score
H-TMB+L-Risk score
L-TMB+H-Risk score
L-TMB+L-Risk score
G
Altered in 39 (88.64%) of 44 ESCC samples.
Altered in 61 (85.92%) of 71 ESCC samples.
TP53
TTN
MUC16
SYNE1
CSMD3
FLG
DNAH5
HMCN1
LRP1B
PCLO
EYS
RYR2
PKHD1L1
SACS
SPTA1
CSMD1
ZNF804B
FAT3
CNTNAP5
NFE2L2
No. of samples
Risk score
Fustat
geneCluster
Missense_Mutation
Nonsense_Mutation
Frame_Shift_Del
In_Frame_Del
Frame_Shift_Ins
Multi_Hit
Riskscore
High
Low
Alive
Dead
Immunity_H
Immunity_L
H
Altered in 56 (94.92%) of 59 EAC samples.
Altered in 56 (93.33%) of 60 EAC samples.

FIGURE 7 |

Tumor Mutational Load of Multi-lncRNA Signatures

To understand the potential role of tumor mutational load in esophageal cancer, somatic mutation data of ESCC and EAC patients were collected and corresponding TMB scores were calculated (**Supplementary Table S12**). The results indicated that in ESCC and EAC patients, high risk always implied high TMB scores. The risk score was positively correlated to the TMB score in EAC ($R = 0.46$, $p < 0.05$), but weak correlation in ESCC ($R = 0.21$, $p < 0.05$) (**Figures 7A–D**). In addition, we identified "High-TMB" and "Low-TMB" groups by a cutoff of the median and performed survival analysis. It revealed that the "TMB-High" group had worse survival than the "TMB-Low" group in ESCC and EAC. Besides, it seemed that there was a combined influence of TMB and multi-lncRNA signature on patient survival outcomes in ESCC and EAC (**Figures 6E,F**). Waterfall plots showed the concrete mutation differences of the top 20 genes between "High-risk" and "Low-risk" groups in ESCC and EAC (**Figures 6G,H**). The same with TMB scores, the mutation frequency in high-risk patients was higher. The top 20 genes of ESCC and EAC were distinct in order, and the mutant genes were consistent. TP53 was the gene with the highest mutation frequency in ESCC and EAC. However, in EAC, the mutation of TP53 was more frequent. Compared to EAC, the frameshift deletion or insertion and in-frame deletion were more common in the top 20 mutant genes of ESCC. The mutations in EAC patients were mainly composed of multi-hits and nonsense mutation.

DISCUSSION

In this study, for a deeper understanding of the tumor microenvironment, we comprehensively analyzed immune infiltration *via* ssGSEA and ESTIMATE algorithms to evaluate the fraction of immune cells in two histopathological subtypes of esophageal cancer ESCC and EAC, respectively. In ESCC and EAC cohorts, the tumor microenvironment was similar to some degree. High-immunity groups always had lower tumor purity and poorer survival. The expression of HLA family members and immune checkpoints was also higher in high-immunity groups. However, there were some differences in TILs, GO terms, and KEGG pathways. In the high-immunity group of ESCC, the amounts of M2 cells would increase characteristically, while in the high-immunity group of EAC, the amounts of activated dendric cells would decline conversely. When the classification criterion turned to risk level. In high risk of ESCC, compared with that of EAC, the amounts of NK cells might decrease. Towards the difference of TILs between ESCC and EAC, it has been widely studied by a lot of researchers. According to the previous studies, the increase of IL-17-releasing mast cells can be a potential prognostic marker and were positively correlated with CD8+T cells and macrophages in the same site in ESCC, indirectly mediating their tumor activity by promoting the recruitment of other effector immune cells (Gong et al., 2019; Han et al., 2020). CD8+T cells in esophageal cancer have been proved to be associated with survival rate, response to neoadjuvant chemotherapy, and lymph node metastasis rate (Zheng et al., 2020). In addition, helper T cells of type 17 (Th17) show contradictory functions in the regulation of esophageal cancer tumor growth. Th17 can promote the invasion of EAC cells but plays a protective role in ESCC by enhancing the cytotoxic effect of natural killer (NK) cells and activating CD1A + DC in tumors (Liu et al., 2017; Melo et al., 2020). Furthermore, tumor-associated macrophages could induce angiogenesis and invasion. Tumor-associated fibroblasts can secrete growth factors, alter the extracellular matrix, and promote tumor migration and metastasis (Kashima et al., 2019). The amounts and activity of TILs are considered to be the key factors to determine the effect of ICB and can predict the prognosis of esophageal cancer (Yagi et al., 2019).

In order to seek for more reasonable prognostic predictors in ESCC and EAC, a prognostic 5-lncRNA signature for ESCC and an 8-lncRNA signature for EAC by Cox regression and Lasso regression analysis, both of which exhibited high accuracy and applicability in predicting the risk of death. We also investigated the clinical traits associated with the risk models of ESCC and EAC. Advanced stages and alcohol consumption were likely correlated with high risk of ESCC patients. Alcohol intake has been testified as an independent risk factor for ESCC currently (Ohashi et al., 2015).

In recent years, studies on lncRNA in esophageal cancer have shown promising results. More and more lncRNAs associated with ESCC or EAC were identified and employed to the diagnosis, prognosis, and therapy. LncRNA CASC9 was considered to upregulate LAMC2 expression and promote ESCC metastasis by interacting with CREB-binding proteins to modify histone acetylation (Liang et al., 2018). Cancer-associated fibroblasts (CAFs) could promote lncRNA DNM3OS to regulate DNA damage reaction, leading to significant radio-resistance (Zhang et al., 2019). LncRNA PVT1A has been proved to serve as a therapeutic target for EAC. Combined targeting of PVT1 and YAP1 might benefit patients with EAC as well. Among the 5-lncRNA signature of ESCC, lncRNA HOXB-AS3 has been found to be abnormally expressed in non-small cell lung cancer, colon cancer, and acute myeloid leukemia (Huang et al., 2017; Huang et al., 2019; Jiang et al., 2020). In the study by Bin et al., TFAP2A-AS1 has been proved to act as a miRNA sponge for miR-933 and regulate the expression of Smad2 (Wu et al., 2020b). In the 8-lncRNA signature, LINC00662 was certified to be upregulated in EAC. It has been proved to accelerate M2 macrophage polarization and hepatocellular carcinoma progression *via* activating Wnt/β-catenin signaling. Its overexpression promoted the occurrence and development of colon cancer by competitively binding with miR-340-5p to regulate CLDN8/IL22 co-expression and activating ERK signaling pathway (Zhou et al., 2019). What is more, LINC00662 is also closely related to gastric cancer, glioma, chordoma, and so on (Liu et al., 2018a; Wu et al., 2020b; Wang et al., 2020). The remaining lncRNAs in the 8-lncRNA signature were associated with lung adenocarcinoma, colon cancer, and so on (Liu et al., 2018b; Cai et al., 2021). The research directed at the lncRNA in the multi-lncRNA signatures of ESCC and EAC was still deficient. Therefore, more studies should be conducted to help explore novel and promising targets for the therapy of esophageal cancer.

Cisplatin, RTKs inhibitors (sunitinib), EGFR inhibitors (erlotinib and gefitinib), HER-2 inhibitor (lapatinib), and

multikinase inhibitor (sorafenib) were widely used and studied in the treatment of ESCC and EAC (Yang et al., 2020b). In ESCC, the IC_{50} of drugs mentioned above all presented significant difference in high- and low-risk patients. Nevertheless, in EAC, cisplatin and sorafenib showed no difference in IC_{50} between high- and low-risk patients. The two multi-lncRNA signatures of ESCC and EAC might help speculate the effectiveness of therapeutic agents and contribute to personalized treatments.

Finally, we evaluated the relationship between TMB and the risk level. The results manifested that both of our multi-lncRNA signatures were positively correlated with the TMB. Among the top 20 mutant genes, TP53 mutated more frequently in EAC patients. As reported, the mutation of TP53 might be the early events in the development of EAC by participating the process of chronic gastroesophageal reflux disease (Guo et al., 2018; Dang and Chai, 2020). Prior to TMB being explored as a biomarker, the focus was on quantitative testing of PD-L1 to identify patients who could benefit most from ICB treatment. It is currently approved as an adjunct diagnostic for pembrolizumab in NSCLC (Reck et al., 2016; Mok et al., 2019). However, single use of PD-L1 expression level has gradually shown poor prediction in ICB treatment response, so TMB was developed as a complementary biomarker. A therapeutic benefit dependent on TMB but independent of PD-L1 expression level was observed in patients treated with a combination of nivolumab + ipilimumab and the standard of care (SOC) chemotherapy (Reck et al., 2019). This condition was thought to be present in tumors with high TMB and T-cell infiltration and/or activation regulated in a CTLA-4-dependent manner (Hu et al., 2020). Moreover, TMB has been found to be of predictive value in immunotherapy other than ICB, with studies showing that TIL therapy has better therapeutic outcomes in patients with higher TMB (Liu et al., 2019; Samstein et al., 2019). Hence, our study for TMB might suggest that high-risk patients would be more responsive to immune therapy on account of the high TMB.

CONCLUSION

In summary, we have comprehensively examined the characteristics of tumor immune microenvironment in esophageal cancer and identified a 5-immune-related lncRNA signature of ESCC and an 8-lncRNA signature of EAC as the prognostic predictor. These two immune-related lncRNA signatures were validated strictly and appeared to be stable. Additional analysis showed that these two multi-lncRNA signatures could be promising biomarkers to predict drug sensitivity as well as benefits from immunotherapy in esophageal cancer based on TMB.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The Cancer Genome Atlas (TCGA, https://tcga-data.nci.nih.gov/tcga/) Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) UCSC Xena (http://xena.ucsc.deu/).

AUTHOR CONTRIBUTIONS

Conceiving and designing the study: JP, HP, CY, PM, and S-YX. Acquisition, analysis, and interpretation of data for the work: JP, HP, WX, YL, and S-YX. Writing the manuscript: JP, HP, and YL. Figure preparation: JP, JL, and CY. Revising the work critically for important intellectual content: PM, YL, and S-YX. Final approval of the version to be published: YL and S-YX. All authors read and approved the final manuscript.

FUNDING

This research was supported by Zhongnan Hospital of Wuhan University Science, Technology and Innovation Seed Fund (znpy2019092).

ACKNOWLEDGMENTS

The authors thank PM for editing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.722601/full#supplementary-material

REFERENCES

Baba, Y., Nomoto, D., Okadome, K., Ishimoto, T., Iwatsuki, M., Miyamoto, Y., et al. (2020). Tumor Immune Microenvironment and Immune Checkpoint Inhibitors in Esophageal Squamous Cell Carcinoma. *Cancer Sci.* 111 (9), 3132–3141. doi:10.1111/cas.14541

Bader, J. E., Voss, K., and Rathmell, J. C. (2020). Targeting Metabolism to Improve the Tumor Microenvironment for Cancer Immunotherapy. *Mol. Cel* 78 (6), 1019–1033. doi:10.1016/j.molcel.2020.05.034

Cai, H.-j., Zhuang, Z.-c., Wu, Y., Zhang, Y.-y., Liu, X., Zhuang, J.-f., et al. (2021). Development and Validation of a Ferroptosis-Related lncRNAs Prognosis Signature in colon Cancer. *Bosn J. Basic Med. Sci.* 21 (5), 569–576. doi:10.17305/bjbms.2020.5617

Dang, T., and Chai, J. (2020). Molecular Dynamics in Esophageal Adenocarcinoma: Who's in Control? *Ccdt* 20 (10), 789–801. doi:10.2174/1568009620666200720011341

Dong, J., Buas, M. F., Gharahkhani, P., Kendall, B. J., Onstad, L., Zhao, S., et al. (2018). Determining Risk of Barrett's Esophagus and Esophageal Adenocarcinoma Based on Epidemiologic Factors and Genetic Variants. *Gastroenterology* 154 (5), 1273–1281e3. doi:10.1053/j.gastro.2017.12.003

Fatehi Hassanabad, A., Chehade, R., Breadner, D., and Raphael, J. (2020). Esophageal Carcinoma: Towards Targeted Therapies. *Cell Oncol.* 43 (2), 195–209. doi:10.1007/s13402-019-00488-2

Gong, W., Hoffmann, J.-M., Stock, S., Wang, L., Liu, Y., Schubert, M.-L., et al. (2019). Comparison of IL-2 vs IL-7/IL-15 for the Generation of NY-ESO-1-specific T Cells. *Cancer Immunol. Immunother.* 68 (7), 1195–1209. doi:10.1007/s00262-019-02354-4

Guo, X., Tang, Y., and Zhu, W. (2018). Distinct Esophageal Adenocarcinoma Molecular Subtype Has Subtype-specific Gene Expression and Mutation Patterns. *BMC Genomics* 19 (1), 769. doi:10.1186/s12864-018-5165-0

Han, L., Gao, Q.-L., Zhou, X.-M., Shi, C., Chen, G.-Y., Song, Y.-P., et al. (2020). Characterization of CD103+ CD8+ Tissue-Resident T Cells in Esophageal Squamous Cell Carcinoma: May Be Tumor Reactive and Resurrected by Anti-PD-1 Blockade. *Cancer Immunol. Immunother.* 69 (8), 1493–1504. doi:10.1007/s00262-020-02562-3

Hinshaw, D. C., and Shevde, L. A. (2019). The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res.* 79 (18), 4557–4566. doi:10.1158/0008-5472.CAN-18-3962

Hu, H., She, L., Liao, M., Shi, Y., Yao, L., Ding, D., et al. (2020). Cost-Effectiveness Analysis of Nivolumab Plus Ipilimumab vs. Chemotherapy as First-Line Therapy in Advanced Non-small Cell Lung Cancer. *Front. Oncol.* 10, 1649. doi:10.3389/fonc.2020.01649

Huang, H.-H., Chen, F.-Y., Chou, W.-C., Hou, H.-A., Ko, B.-S., Lin, C.-T., et al. (2019). Long Non-coding RNA HOXB-AS3 Promotes Myeloid Cell Proliferation and its Higher Expression Is an Adverse Prognostic Marker in Patients with Acute Myeloid Leukemia and Myelodysplastic Syndrome. *BMC Cancer* 19 (1), 617. doi:10.1186/s12885-019-5822-y

Huang, J.-Z., Chen, M., Chen, D., Gao, X.-C., Zhu, S., Huang, H., et al. (2017). A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol. Cel* 68 (1), 171–184e6. doi:10.1016/j.molcel.2017.09.015

Huang, T.-X., and Fu, L. (2019). The Immune Landscape of Esophageal Cancer. *Cancer Commun.* 39 (1), 79. doi:10.1186/s40880-019-0427-z

Jarosz-Biej, M., Smolarczyk, R., Kułach, T., and Kulach, N. (2019). Tumor Microenvironment as A "Game Changer" in Cancer Radiotherapy. *Ijms* 20 (13), 3212. doi:10.3390/ijms20133212

Jiang, W., Kai, J., Li, D., Wei, Z., Wang, Y., and Wang, W. (2020). lncRNA HOXB-AS3 Exacerbates Proliferation, Migration, and Invasion of Lung Cancer via Activating the PI3K-AKT Pathway. *J. Cel. Physiol.* 235 (10), 7194–7203. doi:10.1002/jcp.29618

Kang, K., Xie, F., Mao, J., Bai, Y., and Wang, X. (2020). Significance of Tumor Mutation Burden in Immune Infiltration and Prognosis in Cutaneous Melanoma. *Front. Oncol.* 10, 573141. doi:10.3389/fonc.2020.573141

Kashima, H., Noma, K., Ohara, T., Kato, T., Katsura, Y., Komoto, S., et al. (2019). Cancer-associated Fibroblasts (CAFs) Promote the Lymph Node Metastasis of Esophageal Squamous Cell Carcinoma. *Int. J. Cancer* 144 (4), 828–840. doi:10.1002/ijc.31953

Kobayashi, Y., Kushihara, Y., Saito, N., Yamaguchi, S., and Kakimi, K. (2020). A Novel Scoring Method Based on RNA-Seq Immunograms Describing Individual Cancer-immunity Interactions. *Cancer Sci.* 111 (11), 4031–4040. doi:10.1111/cas.14621

Liang, Y., Chen, X., Wu, Y., Li, J., Zhang, S., Wang, K., et al. (2018). LncRNA CASC9 Promotes Esophageal Squamous Cell Carcinoma Metastasis through Upregulating LAMC2 Expression by Interacting with the CREB-Binding Protein. *Cell Death Differ* 25 (11), 1980–1995. doi:10.1038/s41418-018-0084-9

Liu, A. N., Qu, H. J., Yu, C. Y., and Sun, P. (2018). Knockdown of LINC01614 Inhibits Lung Adenocarcinoma Cell Progression by Up-regulating miR-217 and down-regulatingFOXP1. *J. Cel. Mol. Med.* 22 (9), 4034–4044. doi:10.1111/jcmm.13483

Liu, D., Zhang, R., Wu, J., Pu, Y., Yin, X., Cheng, Y., et al. (2017). Interleukin-17A Promotes Esophageal Adenocarcinoma Cell Invasiveness through ROS-dependent, NF-Kb-Mediated MMP-2/9 Activation. *Oncol. Rep.* 37 (3), 1779–1785. doi:10.3892/or.2017.5426

Liu, L., Bai, X., Wang, J., Tang, X.-R., Wu, D.-H., Du, S.-S., et al. (2019). Combination of TMB and CNA Stratifies Prognostic and Predictive Responses to Immunotherapy across Metastatic Cancer. *Clin. Cancer Res.* 25 (24), 7413–7423. doi:10.1158/1078-0432.CCR-19-0558

Liu, Z., Yao, Y., Huang, S., Li, L., Jiang, B., Guo, H., et al. (2018). LINC00662 Promotes Gastric Cancer Cell Growth by Modulating the Hippo-YAP1 Pathway. *Biochem. Biophysical Res. Commun.* 505 (3), 843–849. doi:10.1016/j.bbrc.2018.09.191

Melo, A. M., Conroy, M. J., Foley, E. K., Dockry, É., Breen, E. P., Reynolds, J. V., et al. (2020). CD1d Expression and Invariant Natural Killer T-Cell Numbers Are Reduced in Patients with Upper Gastrointestinal Cancers and Are Further Impaired by Commonly Used Chemotherapies. *Cancer Immunol. Immunother.* 69 (6), 969–982. doi:10.1007/s00262-020-02514-x

Meurette, O., and Mehlen, P. (2018). Notch Signaling in the Tumor Microenvironment. *Cancer Cell* 34 (4), 536–548. doi:10.1016/j.ccell.2018.07.009

Mok, T. S. K., Wu, Y. L., Kudaba, I., Kowalski, D. M., Cho, B. C., Turna, H. Z., et al. (2019). Pembrolizumab versus Chemotherapy for Previously Untreated, PD-L1-Expressing, Locally Advanced or Metastatic Non-small-cell Lung Cancer (KEYNOTE-042): a Randomised, Open-Label, Controlled, Phase 3 Trial. *Lancet* 393 (10183), 1819–1830. doi:10.1016/S0140-6736(18)32409-7

Ohashi, S., Miyamoto, S. i., Kikuchi, O., Goto, T., Amanuma, Y., and Muto, M. (2015). Recent Advances from Basic and Clinical Studies of Esophageal Squamous Cell Carcinoma. *Gastroenterology* 149 (7), 1700–1715. doi:10.1053/j.gastro.2015.08.054

Qian, X., Zhao, J., Yeung, P. Y., Zhang, Q. C., and Kwok, C. K. (2019). Revealing lncRNA Structures and Interactions by Sequencing-Based Approaches. *Trends Biochem. Sci.* 44 (1), 33–52. doi:10.1016/j.tibs.2018.09.012

Reck, M., Rodríguez-Abreu, D., Robinson, A. G., Hui, R., Csőszi, T., Fülöp, A., et al. (2016). Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-small-cell Lung Cancer. *N. Engl. J. Med.* 375 (19), 1823–1833. doi:10.1056/NEJMoa1606774

Reck, M., Schenker, M., Lee, K. H., Provencio, M., Nishio, M., Lesniewski-Kmak, K., et al. (2019). Nivolumab Plus Ipilimumab versus Chemotherapy as First-Line Treatment in Advanced Non-small-cell Lung Cancer with High Tumour Mutational burden: Patient-Reported Outcomes Results from the Randomised, Open-Label, Phase III CheckMate 227 Trial. *Eur. J. Cancer* 116, 137–147. doi:10.1016/j.ejca.2019.05.008

Riley, R. S., June, C. H., Langer, R., and Mitchell, M. J. (2019). Delivery Technologies for Cancer Immunotherapy. *Nat. Rev. Drug Discov.* 18 (3), 175–196. doi:10.1038/s41573-018-0006-z

Robinson, E. K., Covarrubias, S., and Carpenter, S. (2020). The How and Why of lncRNA Function: An Innate Immune Perspective. *Biochim. Biophys. Acta (Bba) - Gene Regul. Mech.* 1863 (4), 194419. doi:10.1016/j.bbagrm.2019.194419

Samstein, R. M., Lee, C.-H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., et al. (2019). Tumor Mutational Load Predicts Survival after Immunotherapy across Multiple Cancer Types. *Nat. Genet.* 51 (2), 202–206. doi:10.1038/s41588-018-0312-8

Shitara, K., Özgüroğlu, M., Bang, Y. J., Di Bartolomeo, M., Mandalà, M., Ryu, M. H., et al. (2018). Pembrolizumab versus Paclitaxel for Previously Treated, Advanced Gastric or Gastro-Oesophageal junction Cancer (KEYNOTE-061): a Randomised, Open-Label, Controlled, Phase 3 Trial. *Lancet* 392 (10142), 123–133. doi:10.1016/S0140-6736(18)31257-1

Song, C., Guo, Z., Yu, D., Wang, Y., Wang, Q., Dong, Z., et al. (2020). A Prognostic Nomogram Combining Immune-Related Gene Signature and Clinical Factors Predicts Survival in Patients with Lung Adenocarcinoma. *Front. Oncol.* 10, 1300. doi:10.3389/fonc.2020.01300

Talukdar, F. R., di Pietro, M., Secrier, M., Moehler, M., Goepfert, K., Lima, S. S. C., et al. (2018). Molecular Landscape of Esophageal Cancer: Implications for Early Detection and Personalized Therapy. *Ann. N.Y. Acad. Sci.* 1434 (1), 342–359. doi:10.1111/nyas.13876

Thrift, A. P. (2021). Global burden and Epidemiology of Barrett Oesophagus and Oesophageal Cancer. *Nat. Rev. Gastroenterol. Hepatol.* 18 (6), 432–443. doi:10.1038/s41575-021-00419-3

Wang, C. B., Wang, Y., Wang, J. J., and Guo, X. L. (2020). LINC00662 Triggers Malignant Progression of Chordoma by the Activation of RNF144B via Targeting miR-16-5p. *Eur. Rev. Med. Pharmacol. Sci.* 24 (3), 1007–1022. doi:10.26355/eurrev_202002_20151

Wang, M., Xu, T., Feng, W., Liu, J., and Wang, Z. (2021). Advances in Understanding the LncRNA-Mediated Regulation of the Hippo Pathway in Cancer. *Ott* 14, 2397–2415. doi:10.2147/OTT.S283157

Wu, J., Guo, X., Xu, D., and Zhang, H. (2020). LINC00662 Sponges miR-107 Accelerating the Invasiveness and Proliferation of Glioma Cells. *J. Cancer* 11 (19), 5700–5712. doi:10.7150/jca.46381

Wu, T., and Dai, Y. (2017). Tumor Microenvironment and Therapeutic Response. *Cancer Lett.* 387, 61–68. doi:10.1016/j.canlet.2016.01.043

Wu, Y., Sang, M., Liu, F., Zhang, J., Li, W., Li, Z., et al. (2020). Epigenetic Modulation Combined with PD-1/pd-L1 Blockade Enhances Immunotherapy Based on MAGE-A11 Antigen-specific CD8+T Cells against Esophageal Carcinoma. *Carcinogenesis* 41 (7), 894–903. doi:10.1093/carcin/bgaa057

Yagi, T., Baba, Y., Ishimoto, T., Iwatsuki, M., Miyamoto, Y., Yoshida, N., et al. (2019). PD-L1 Expression, Tumor-Infiltrating Lymphocytes, and Clinical Outcome in Patients with Surgically Resected Esophageal Cancer. *Ann. Surg.* 269 (3), 471–478. doi:10.1097/SLA.0000000000002616

Yang, H., Wang, K., Wang, T., Li, M., Li, B., Li, S., et al. (2020). The Combination Options and Predictive Biomarkers of PD-1/pd-L1 Inhibitors in Esophageal Cancer. *Front. Oncol.* 10, 300. doi:10.3389/fonc.2020.00300

Yang, L., Yang, Y., Meng, M., Wang, W., He, S., Zhao, Y., et al. (2021). Identification of Prognosis-Related Genes in the Cervical Cancer Immune Microenvironment. *Gene* 766, 145119. doi:10.1016/j.gene.2020.145119

Yang, Y.-M., Hong, P., Xu, W. W., He, Q.-Y., and Li, B. (2020). Advances in Targeted Therapy for Esophageal Cancer. *Sig Transduct Target. Ther.* 5 (1), 229. doi:10.1038/s41392-020-00323-3

Zhang, H., Hua, Y., Jiang, Z., Yue, J., Shi, M., Zhen, X., et al. (2019). Cancer-associated Fibroblast-Promoted LncRNA DNM3OS Confers Radioresistance by Regulating DNA Damage Response in Esophageal Squamous Cell Carcinoma. *Clin. Cancer Res.* 25 (6), 1989–2000. doi:10.1158/1078-0432.CCR-18-0773

Zheng, Y., Chen, Z., Han, Y., Han, L., Zou, X., Zhou, B., et al. (2020). Immune Suppressive Landscape in the Human Esophageal Squamous Cell Carcinoma Microenvironment. *Nat. Commun.* 11 (1), 6268. doi:10.1038/s41467-020-20019-0

Zhou, B., Guo, H., and Tang, J. (2019). Long Non-coding RNA TFAP2A-AS1 Inhibits Cell Proliferation and Invasion in Breast Cancer via miR-933/SMAD2. *Med. Sci. Monit.* 25, 1242–1253. doi:10.12659/MSM.912421

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

GLOSSARY

APC antigen-presenting cells

AUC area under curve

CAF cancer-associated fibroblasts

CAR-T chimeric antigen receptor T cells

CRT chemoradiotherapy

CSC cancer stem cell

DC dendritic cells

EAC esophageal adenocarcinoma

EGFR epidermal growth factor receptor

ESTIMATE Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression

FDA Food and Drug Administration

GO Gene Oncology

HER-2 human epidermal growth factor receptor-2

HLAs human leukocyte antigens

ICB immune-checkpoint blockade

KEGG Kyoto Encyclopedia of Genes and Genomes

LASSO Least Absolute Shrinkage and Selection Operator

LncRNA long non-coding RNA

MDSCs myeloid derived suppressor cells

MHC major histocompatibility complex

NK natural killer

NSCLC non-small cell lung cancer

PCA principal components analysis

RESCC esophageal squamous cell carcinoma

ROC receiver operating characteristic

RTKs receptor tyrosine kinases

SOC standard of care

ssGSEA single sample Gene Set Enrichment Analysis

TCGA The Cancer Genome AtlasThe Cancer Genome Atlas

TCGA The Cancer Genome AtlasThe Cancer Genome Atlas

Th17 helper T cells of type 17

TILs tumor-infiltrating lymphocytes

TMB tumor mutation burden

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 November 2021
doi: 10.3389/fgene.2021.791640

HIF-Dependent *NFATC1* Activation Upregulates *ITGA5* and *PLAUR* in Intestinal Epithelium in Inflammatory Bowel Disease

Evgeny Knyazev[1]*, *Diana Maltseva*[2], *Maria Raygorodskaya*[1] and *Maxim Shkurnikov*[1,2,3]

[1]Laboratory of Microfluidic Technologies for Biomedicine, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Moscow, Russia, [2]Faculty of Biology and Biotechnology, National Research University Higher School of Economics (HSE), Moscow, Russia, [3]National Center of Medical Radiological Research, P. Hertsen Moscow Oncology Research Institute, Moscow, Russia

OPEN ACCESS

Edited by:
Yuriy L. Orlov,
I.M. Sechenov First Moscow State Medical University, Russia

Reviewed by:
Andrey Zamyatnin,
I.M. Sechenov First Moscow State Medical University, Russia
Kerim Mutig,
Charité University Medicine Berlin, Germany
Vladimir S. Prassolov,
Engelhardt Institute of Molecular Biology (RAS), Russia

***Correspondence:**
Evgeny Knyazev
knyazevevg@gmail.com

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 08 October 2021
Accepted: 27 October 2021
Published: 11 November 2021

Citation:
Knyazev E, Maltseva D, Raygorodskaya M and Shkurnikov M (2021) HIF-Dependent NFATC1 Activation Upregulates ITGA5 and PLAUR in Intestinal Epithelium in Inflammatory Bowel Disease. Front. Genet. 12:791640. doi: 10.3389/fgene.2021.791640

Intestinal epithelial cells exist in physiological hypoxia, leading to hypoxia-inducible factor (HIF) activation and supporting barrier function and cell metabolism of the intestinal epithelium. In contrast, pathological hypoxia is a common feature of some chronic disorders, including inflammatory bowel disease (IBD). This work was aimed at studying HIF-associated changes in the intestinal epithelium in IBD. In the first step, a list of genes responding to chemical activation of hypoxia was obtained in an *in vitro* intestinal cell model with RNA sequencing. Cobalt (II) chloride and oxyquinoline treatment of both undifferentiated and differentiated Caco-2 cells activate the HIF-signaling pathway according to gene set enrichment analysis. The core gene set responding to chemical hypoxia stimulation in the intestinal model included 115 upregulated and 69 downregulated genes. Of this set, protein product was detected for 32 genes, and fold changes in proteome and RNA sequencing significantly correlate. Analysis of publicly available RNA sequencing set of the intestinal epithelial cells of patients with IBD confirmed HIF-1 signaling pathway activation in sigmoid colon of patients with ulcerative colitis and terminal ileum of patients with Crohn's disease. Of the core gene set from the gut hypoxia model, expression activation of ITGA5 and PLAUR genes encoding integrin α5 and urokinase-type plasminogen activator receptor (uPAR) was detected in IBD specimens. The interaction of these molecules can activate cell migration and regenerative processes in the epithelium. Transcription factor analysis with the previously developed miRGTF tool revealed the possible role of HIF1A and NFATC1 in the regulation of ITGA5 and PLAUR gene expression. Detected genes can serve as markers of IBD progression and intestinal hypoxia.

Keywords: intestinal bowel disease, hypoxia, cobalt, hydroxyquinolines, caco-2 cells, urokinase-type plasminogen activator receptor, disease markers, transcriptomics

1 INTRODUCTION

Some pathologic conditions, such as inflammatory bowel disease (IBD), including ulcerative colitis (UC) and Crohn's disease (CD), are associated with pathological hypoxia (Taylor, 2018). UC affects only the colon, while CD affects both the small and large intestines. The exact pathogenesis of IBD is still unclear, but it is a complex pathology involving genetics, environment, microbiome, and immunome (Borg-Bartolo et al., 2020). Hypoxia leading to hypoxia-inducible factor (HIF)-signaling pathway activation is a common feature of inflammatory diseases, including IBD (Kerber et al., 2020). HIF-1α is protective in IBD models (Kim et al., 2021), and HIF-2α is essential in maintaining the immune response and regenerative capacity of the intestinal epithelium in IBD (Ramakrishnan and Shah, 2016). At the same time, HIF-2α but not HIF-1α constitutive activation leads to IBD development or exacerbation in colitis models (Xue et al., 2013; Solanki et al., 2019). HIF-1α activation in myeloid cells aggravates, and HIF-2α activation ameliorated IBD in murine models (Kim et al., 2018; Kerber et al., 2020). These results emphasize that the degree of hypoxia and HIF activation can lead to both adaptation and damage to the intestinal epithelium.

The human small intestine consists of crypts and villi, and the colon consists of crypts only. Crypts are invaginations aligned by intestinal stem cells, early progenitor transit-amplifying cells, and differentiating cells (Rangel-Huerta and Maldonado, 2017). The colon surface and small intestine villi are covered by specialized epithelium with differentiated cells, such as enterocytes and goblet cells (Singhal and Shah, 2020). The main enterocyte functions are to create a barrier between the external and internal environment and to provide regulated transport of substances between these environments (Schoultz and Keita, 2020). There is a unique oxygen gradient from the crypt to the intestinal lumen. Electron paramagnetic resonance oximetry revealed partial oxygen pressure (pO_2) gradient from 59 mm Hg (8%) in the small intestine wall to 22 mm Hg (3%) at the villus apex and <10 mm Hg (2%) in the small intestinal lumen, whereas colon pO_2 is 5–10 mm Hg near the crypt-lumen interface, and 11 (~2%) and 3 mm Hg (~0.4%) in the lumen of ascending and sigmoid colon, respectively (Singhal and Shah, 2020). Thus, intestinal cells exist under conditions ranging from mild to pronounced hypoxia, which leads to HIF stabilization and adaptation to hypoxic conditions (Pral et al., 2021). HIF stabilization in physiological hypoxia supports cell metabolism and barrier function of the intestinal epithelium (Glover et al., 2016).

Modeling of the human intestinal epithelium is possible with human colon adenocarcinoma cell line Caco-2 (Nikulin et al., 2018b). Undifferentiated Caco-2 (uCaco-2) cells actively proliferate, resembling intestinal stem cells, but contact inhibition stimulates differentiation and suppresses proliferation. Differentiated Caco-2 (dCaco-2) cells acquire enterocyte properties, such as cylindrical cell shape, formation of intercellular junctions and microvilli, specific enzyme and marker expression (Ding et al., 2021). Extracellular matrix can support cell attachment, proliferation, and differentiation (Mutsenko et al., 2017), and use of laminins specific for the intestinal basal membrane can bring the gut model properties closer to conditions *in vivo* (Nikulin et al., 2018a, 2019). Microfluidic devices with a culture medium flow that simulates blood flow create even more physiological conditions for the cell model (Samatov et al., 2016, 2017; Sakharov et al., 2019). Caco-2 intestinal model is suitable to study hypoxia effects (Maltseva et al., 2020; Nersisyan et al., 2021a). Small molecules can fit in the enzyme active center inhibiting its activity (Smirnov et al., 2011; Ivanenkov et al., 2015). Oxyquinoline derivatives can directly block the prolyl hydroxylase active center leading to HIF stabilization (Knyazev et al., 2019a; 2019b). Cobalt (II) chloride ($CoCl_2$) also serves as a chemical hypoxia inductor by replacing Fe^{2+} ions in the prolyl hydroxylase active center, induction of ascorbate oxidation, and other potential mechanisms (Muñoz-Sánchez and Chánez-Cárdenas, 2019). Previously, we showed that chemical hypoxia models, induced by $CoCl_2$ and OD, may serve as models of severe and mild hypoxia, respectively (Knyazev et al., 2021; Knyazev and Paul, 2021).

This work aimed to study the effects of chemical HIF activation on dCaco-2 and uCaco-2 cells as the intestinal model and compare observed changes with patterns in the intestinal epithelium of IBD patients to find hypoxic markers, relevant for IBD.

2 MATERIALS AND METHODS

2.1 Cell Cultures and Treatments

Caco-2 cells were received from the Russian Vertebrate Cell Culture Collection (Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia). The cells were incubated in Gibco minimal essential medium with L-glutamine, 20% Gibco FBS One Shot, and 1% Gibco Pen Strep (Thermo Fisher Scientific, United States). The cells were seeded in the 6-well culture plates (TPP Techno Plastic Products AG, Germany) at a seeding density of 0.3 x 10^6 cells per well and cultivated for 21 days to achieve cell differentiation, as previously published (Nersisyan et al., 2021a). On the last week of differentiation new portion of cells was seeded in the 6-well culture plates at the same seeding density and cultivated to 80% of confluence. The cell culture medium of differentiated and undifferentiated Caco-2 was replaced with either medium with 300 μM $CoCl_2$, medium with 5 μM OD 4896–3,212 (ChemDiv Research Institute, Khimki, Russia), or fresh medium. OD stock solution was diluted in DMSO at 10 mM, so 0.5% DMSO was also added to $CoCl_2$ and control medium, and cells were incubated 24 h before lysis for RNA and protein extraction.

2.2 RNA Extraction, Library Preparation and Sequencing

Cells were lysed in 700 μL of QIAzol Lysis Reagent (Qiagen, Hilden, Germany) with subsequent RNA extraction with miRNeasy Mini Kit (Qiagen) according to the manufacturer protocol, including on-column treatment with RNase-Free DNase Set (Qiagen). RNA was extracted in 30 μL of RNase-

free water. RNA concentration was measured with NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific). A_{260}/A_{280} and A_{260}/A_{230} ratios were more than 2.0, and RNA Integrity Number (RIN) according to 2,100 Bioanalyzer with Agilent RNA 6000 Pico Kit (Agilent Technologies, Santa Clara, CA, United States) was not less than 9.7 for all RNA samples.

Libraries for mRNA sequencing were prepared with TruSeq Stranded mRNA Library Prep (Illumina, San Diego, CA, United States) with 1 μg total RNA in three biological replicates for each condition. Library quality check was provided with Agilent High Sensitivity DNA Kit (Agilent Technologies). Sequencing with NextSeq 550 (Illumina) generated single-end 75-nucleotide reads. The sequencing data is available online in the Gene Expression Omnibus (GEO) with the accession numbers GSE186295 and GSE158632.

2.3 Sequencing Data Processing

The FASTQ file quality was assessed with FastQC v0.11.9 (Babraham Bioinformatics, Cambridge, United Kingdom). After adapter trimming with cutadapt v2.10 (Martin, 2011) reads were mapped on the reference human genome (GENCODE GRCh38.p13) with STAR v2.7.5b (Dobin et al., 2013). The count matrix was generated with GENCODE genome annotation (release 34) (Frankish et al., 2019). FASTQ files for intestinal cells in IBD and healthy control were downloaded from the online repository (E-MTAB-5464) (Howell et al., 2018). Differential expression analysis was performed with DESeq2 v1.28.1 (Love et al., 2014); false discovery rates (FDRs) were calculated by the Benjamini–Hochberg procedure. Differences were considered significant at FDR <0.05 and log fold changes modulo >1.0.

2.4 Gene Set Enrichment Analysis and Transcription Factors Search

Gene set processing and Venn diagram creation was made with jvenn tool (Bardou et al., 2014). Gene set enrichment analysis was performed using Fgsea package version 1.16.0 (Korotkevich et al., 2016). Gene sets were downloaded from Gene Set Enrichment Analysis (GSEA) site (http://www.gsea-msigdb.org), from The Molecular Signatures database (MSigDB 7.4) to assess the activity of hallmark gene sets (Subramanian et al., 2005; Liberzon et al., 2015). The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Huang et al., 2009) was used to reveal Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways activity (Kanehisa et al., 2021).

The previously developed miRGTF-net tool was used to find transcription factors regulating genes of interest (Nersisyan et al., 2021b). The publicly available datasets of RNA sequencing data for colorectal cancer were received from The Cancer Genome Atlas Program Colon Adenocarcinoma (TCGA-COAD) cohort (Cancer Genome Atlas Network, 2012).

2.5 Proteome

Differentiated Caco-2 cells after 24 h incubation with $CoCl_2$ and control Caco-2 cells were lysed with ice-cold lysis buffer with 4% SDS and 0.1 M DTT in 0.1 M Tris-HCl (pH 7.6) and briefly sonicated on the ice with CPX 130 Ultrasonic Processor, 130 W, 20 kHz (Cole-Parmer Instruments, Vernon Hills, IL, United States), 30 s in pulse regimen and 30% amplitude. Total protein concentration was measured with Pierce BCA Protein Assay Kit - Reducing Agent Compatible (Thermo Fisher Scientific). Tripsinized protein samples were analyzed with the Q Exactive HF hybrid quadrupole-orbitrap mass spectrometer with nano-electrospray ionization (nESI) source operated in the positive ionization mode (Thermo Fisher Scientific), the emitter voltage of 2.1 kV, the capillary temperature of 240 C. Progenesis IQ software (Waters Corporation, Milford, MA, United States) was used to quantify protein levels with subsequent analysis with the SearchGUI v.3.3.1 software and the HumanDB database (UniProt Release 2018_05). Differential expression was assessed with the iBAQ algorithm using MaxQuant 1.6 software (Max Planck Institute of Biochemistry, Martinsried, Germany). Differences were considered significant at a fold change >2.0 and Student's t-test $p < 0.05$.

3 RESULTS

3.1 General Pattern of Expression Changes in Response to Cobalt (II) Chloride and Oxyquinoline Derivative Exposure in Differentiated and Undifferentiated Caco-2 Cells

There is no generally accepted list of genes that should respond to HIF activation in all cells and tissues, although some studies show common patterns (Benita et al., 2009). We performed an analysis of gene expression changes in dCaco-2 and uCaco-2 upon exposure to OD and $CoCl_2$ as known activators of the HIF-signaling pathway. The total number of genes whose expression changed significantly in at least one experimental group was 7,180. The pattern of expression changes was similar in all treatments with a notably stronger effect of $CoCl_2$ in both cell types, the most pronounced in uCaco-2 (**Figure 1** and **Supplementary Table S1**).

We performed gene set enrichment analysis to assess the activity of pathways and biological processes in Caco-2 cells upon HIF-pathway activation. We used hallmark gene sets from The MSigDB. Both OD and $CoCl_2$ have activated HALLMARK_HYPOXIA Pathway in uCaco-2 and dCaco-2, confirming the activation of the HIF signaling pathway. Also, in all experimental settings except uCaco-2 OD stimulation, there was activation of the following pathways: HALLMARK_TNFA_SIGNALING_VIA_NFKB, HALLMARK_P53_PATHWAY, and HALLMARK_MTORC1_SIGNALING. It appears that activation of genes regulated by NF-kB in response to tumor necrosis factor α (TNFα) may indicate intersections between hypoxic and inflammatory effects of chemical HIF activators. $CoCl_2$ and OD may activate apoptotic processes through genes involved in the p53 pathway and proliferation through genes upregulated through activation of the mTORC1 complex.

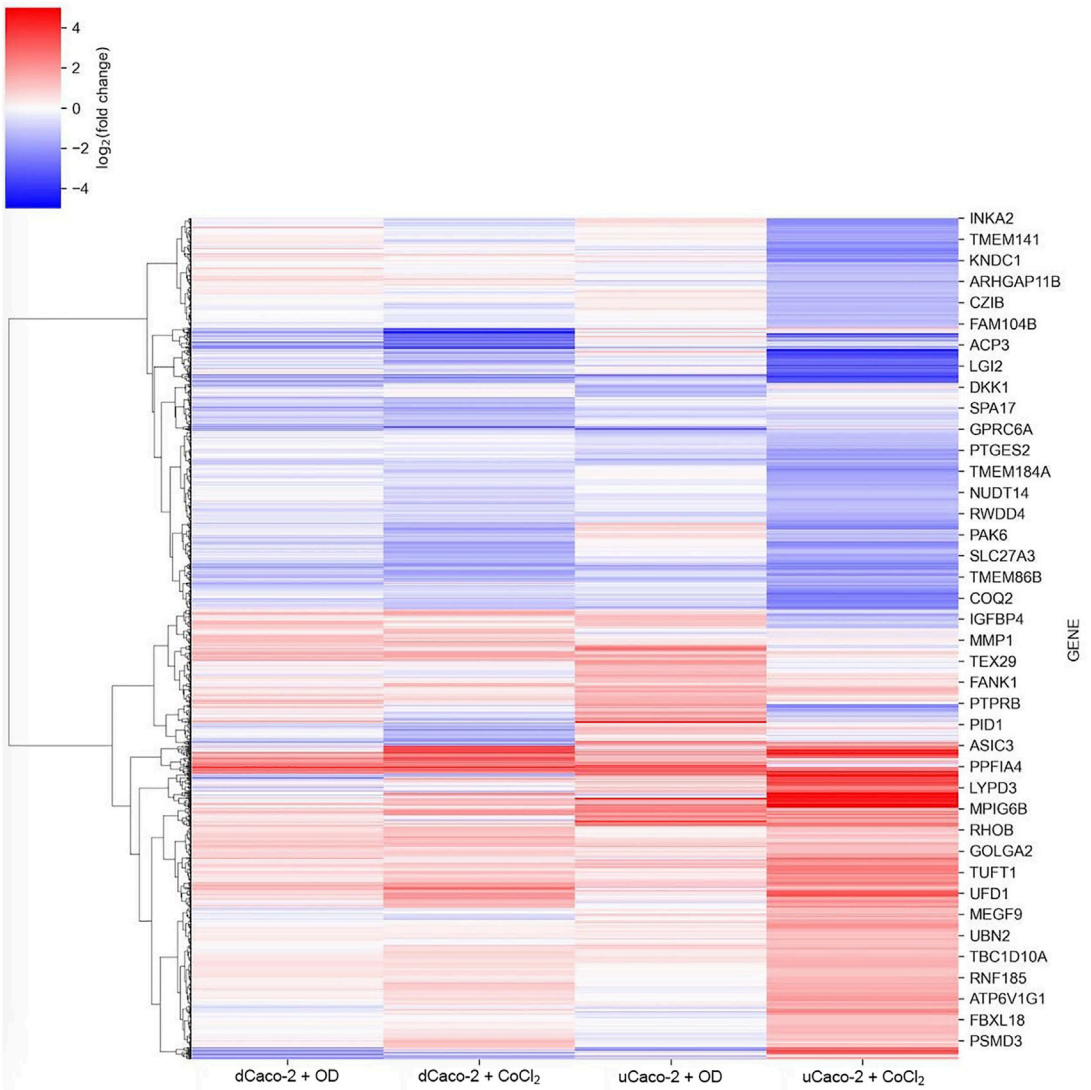


FIGURE 1 | mRNA expression changes for differentiated Caco-2 (dCaco-2) and undifferentiated Caco-2 (uCaco-2) upon cobalt (II) chloride ($CoCl_2$) and oxyquinoline derivative (OD) treatment. The heatmap includes genes with differential expression in at least one combination of cell line and treatment.

$CoCl_2$ specifically activates the following pathways in both uCaco-2 and dCaco-2: HALLMARK_COMPLEMENT, HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY, HALLMARK_INFLAMMATORY_RESPONSE. It indicates activation of inflammatory response in Caco-2 cells, and reactive oxygen species can be generated in cells upon $CoCl_2$ stimulation as a side effect apart from HIF stabilization (Muñoz-Sánchez and Chánez-Cárdenas, 2019).

3.2 Core Gene Set Responding to Chemical Hypoxia Simulation in Caco-2 Cells

The core gene set with the same expression change direction both for uCaco-2 and dCaco-2 upon $CoCl_2$ and OD treatment included 184 genes, 115 upregulated and 69 downregulated (**Figure 2** and **Supplementary Table S2**). As for the complete list of differentially expressed genes, more pronounced expression changes were observed upon $CoCl_2$ exposure, especially in uCaco-2 (**Figure 3**).

Enrichment Analysis of differentially expressed genes in both types of cells and both treatments using DAVID v6.8 revealed significant activation of the following KEGG pathways: Glycolysis/Gluconeogenesis (hsa00010), HIF-1 signaling pathway (hsa04066), Fructose and mannose metabolism (hsa00051), and Carbon metabolism (hsa01200), which is evidence in favor of an activation of the hypoxia signaling pathway and corresponding metabolic changes. This gene set was used as a possible HIF-associated pattern in intestinal cells exposed to hypoxia.

Proteomic profiling of dCaco-2 upon $CoCl_2$ stimulation and in control conditions revealed 3,361 proteins, of them 209 were upregulated and 28 downregulated significantly ($p < 0.05$) after $CoCl_2$ stimulation. Gene set enrichment analysis for these proteins

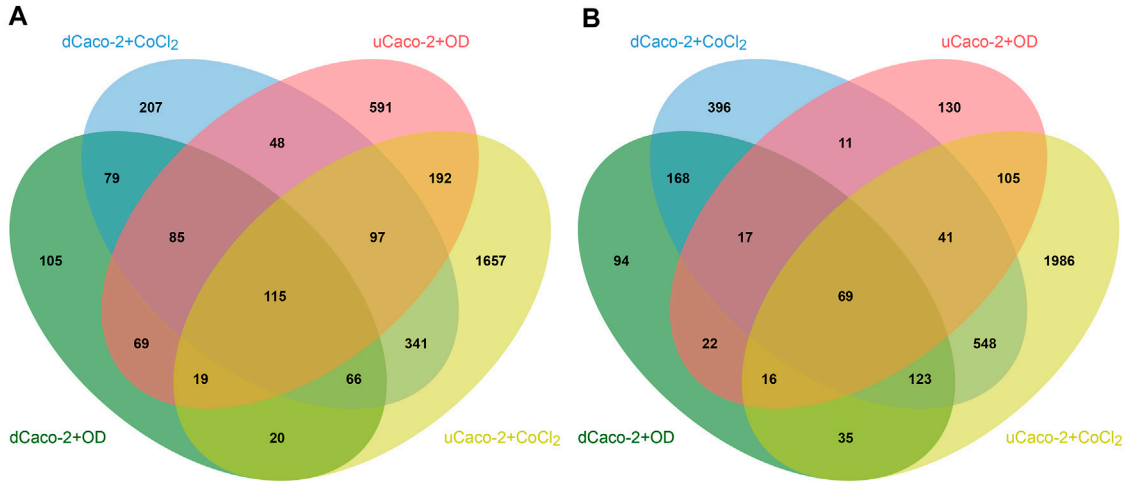


FIGURE 2 | Intersections of differentially expressed genes in differentiated Caco-2 (dCaco-2) and undifferentiated Caco-2 (uCaco-2) upon cobalt (II) chloride ($CoCl_2$) and oxyquinoline derivative (OD) treatment. **(A)**—upregulated genes. **(B)**—downregulated genes.

have not revealed significantly activated pathways at FDR<0.05, but $p < 0.05$ without Benjamini-Hochberg adjustment was revealed for Ribosome (hsa03010) and Endocytosis (hsa04144) KEGG pathways, suggesting an effect of HIF activation on protein synthesis and nutrient transport. Of the 184 genes on the list of HIF-associated genes, a protein product was detected for 32 genes. Analysis of fold changes in proteomic and sequencing data confirmed a significant correlation between these indicators (**Figure 4**). Next, we compared this gene list with specific expression changes in the intestinal epithelium of IBD patients.

3.3 Potential HIF-Associated Changes in Intestinal Epithelium of Patients With Inflammatory Bowel Disease

IBD includes CD, which affects both the small and large intestine, and UC, which affects the colon only. We performed an analysis of a publicly available RNA sequencing set of purified intestinal epithelial cells from pediatric biopsies including IBD and healthy controls (E-MTAB-5464) (Howell et al., 2018). We used the data on RNA expression in sigmoid colon of patients with UC and terminal ileum and sigmoid colon of patients with CD and individuals without IBD. Comparison of disease state with healthy control revealed the upregulation of 172 genes in sigmoid colon in CD, 476 genes in terminal ileum in CD, and 922 genes in sigmoid colon in UC. The downregulation was revealed for 7 genes in sigmoid colon in CD, 258 genes in terminal ileum in CD, and 42 genes in sigmoid colon in UC (**Figure 5**).

Gene Set Enrichment Analysis has shown activation of HALLMARK_INFLAMMATORY_RESPONSE in all groups and downregulation of HALLMARK_OXIDATIVE_PHOSPHORYLATION in sigmoid colon in UC and terminal ileum in CD, possibly indicating hypoxic response (Samanta and Semenza, 2017). The common gene set for all pathologies includes 77 genes with significant activation of the following KEGG pathways: Cytokine-cytokine receptor interaction (hsa04060) and Chemokine signaling pathway (hsa04062), which indicates the immune nature of IBD. Moreover, enrichment analysis revealed significant activation of the HIF signaling pathway in terminal ileum in CD and sigmoid colon in UC.

We analyzed the intersection of these IBD-associated genes with possible HIF-associated gene set from the experiment with chemical hypoxia models in Caco-2 cells. We noted upregulation of *ITGA5*, *MICB*, *PLAUR*, and *DYSF* and downregulation of *GSTA2*, *SLC2A2*, and *KDM8* in terminal ileum of patients with CD. Only two genes from Caco-2 experiment-derived gene set were also upregulated in sigmoid colon of patients with CD (*BHLHE40* and *PLAUR*), and eleven genes from this gene set were upregulated in sigmoid colon of patients with UC (*ITGA5, RNF183, BHLHE40, PLAUR, PFKFB3, FSCN1, SOCS3, RNF24, CSRP2, PIK3CD, DYSF*).

The only differentially expressed gene in all samples was *PLAUR*, which encodes the urokinase-type plasminogen activator receptor (uPAR) which may indicate its involvement in the course of IBD. The HIF signaling pathway according to gene set enrichment analysis was activated in terminal ileum in CD and sigmoid colon in UC, and their gene list share *ITGA5*, *PLAUR*, and *DYSF* genes.

3.4 Transcription Factors Regulating *ITGA5* and *PLAUR* Expression

As *ITGA5* and *PLAUR* genes are both upregulated in hypoxia model and IBD specimens and interact with each other to activate intracellular signaling (Smith and Marshall, 2010). We performed an analysis of potential transcription factors regulating both these genes with the previously developed tool miRGTF-net (Nersisyan et al., 2021b) and received 345 potential transcription regulators of *ITGA5* and *PLAUR*, including HIF1A and HIF3A. Next, we analyzed expression correlation for these transcription factors, *ITGA5*, and *PLAUR* in the mRNA-sequencing data of colorectal cancer samples from TCGA to find regulators which expression significantly correlates with both *ITGA5* and *PLAUR* in the

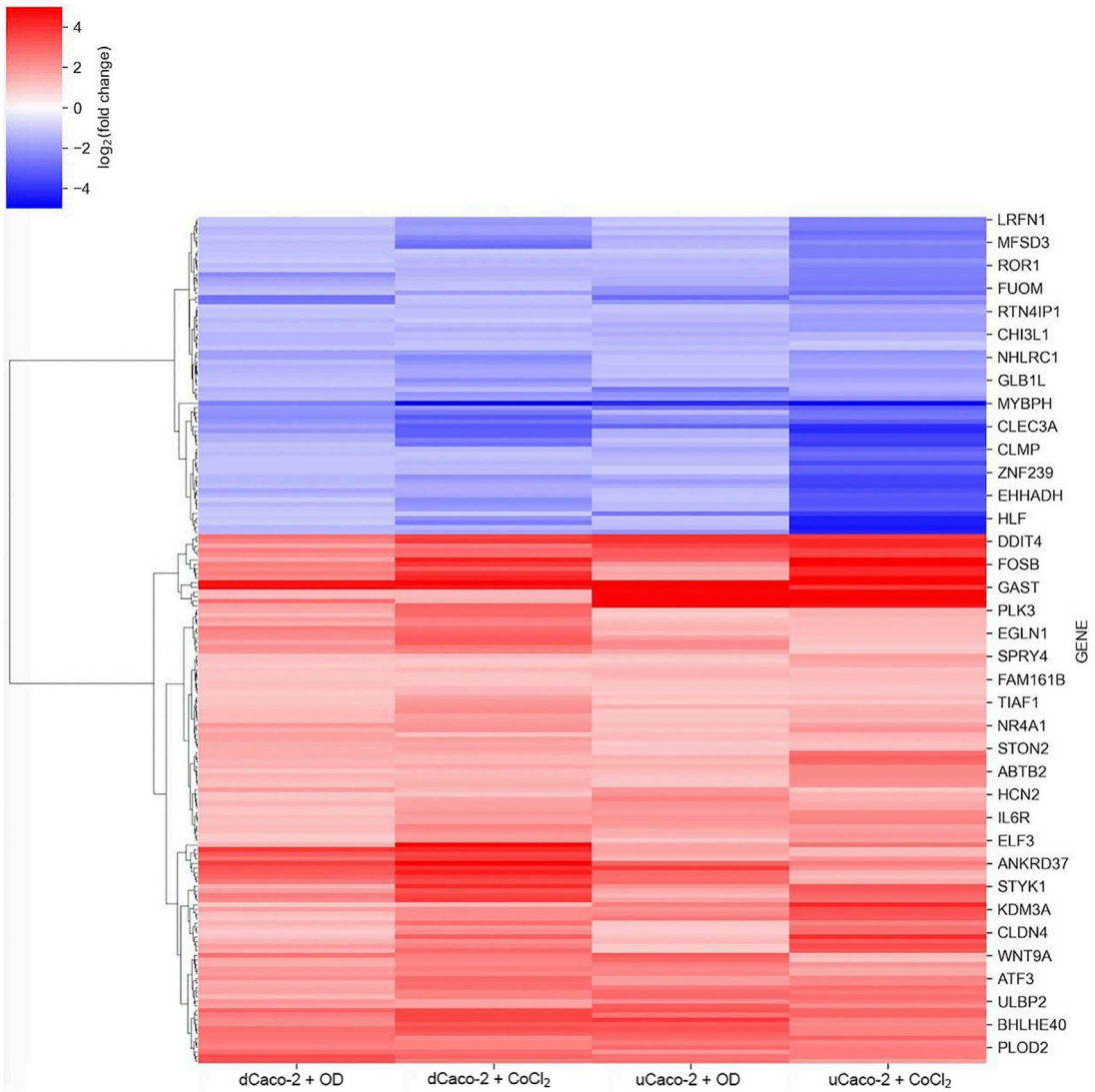


FIGURE 3 | mRNA expression changes for the core gene set reacting to chemical hypoxia simulation in differentiated Caco-2 (dCaco-2) and undifferentiated Caco-2 (uCaco-2) upon cobalt (II) chloride (CoCl2) and oxyquinoline derivative (OD) treatment.

samples of intestinal origin. *GRHL2* and *ZC3H8* have a negative correlation and *MSC*, *NFATC1*, *PRDM1*, and *SPI1* have a positive correlation with *ITGA5* and *PLAUR*. Of these transcription factors, all except SPI1 are regulated by HIF1A according to miRGTF-net. Only NFATC1 has a tendency to be upregulated in Caco-2 according to our RNA sequencing data: 1.6 times in dCaco-2 with OD and uCaco-2 with $CoCl_2$, 1.9 times in dCaco-2 with $CoCl_2$, and 2.9 times in uCaco-2 with OD, suggesting its possible role in *ITGA5* and *PLAUR* upregulation in hypoxia and IBD.

4 DISCUSSION

There is no generally accepted list of genes that always reacting to HIF-pathway activation. Benita et al. have identified HIF-1-target genes combining genomic data from different experiments and found out core gene set of 17 genes that responded to hypoxia in all studied cell types (Benita et al., 2009). Of them, all genes responded to chemical HIF-stabilization at least in one type of cells and treatment options in our experiments, but only nine genes responded in both uCaco-2 and dCaco-2 upon $CoCl_2$ and OD stimulation: *ANKRD37*, *NDRG1*, *PDK1*, *BNIP3*, *DDIT4*, *P4HA1*, *KDM3A*, *BHLHE40*, and *ALDOC*. Only nine genes from this core gene set were detected in proteomic data of dCaco-2 upon $CoCl_2$ stimulation and five of them were differentially expressed on the protein level: NDRG1, PDK1, BNIP3, P4HA1, and ALDOC. Not all genes reacting to hypoxia have a HIF-binding site and some genes respond to HIF stabilization in specific cell sets (Benita et al., 2009). Hypoxia can induce gene expression not only directly through HIF, but also indirectly through regulation of microRNAs and

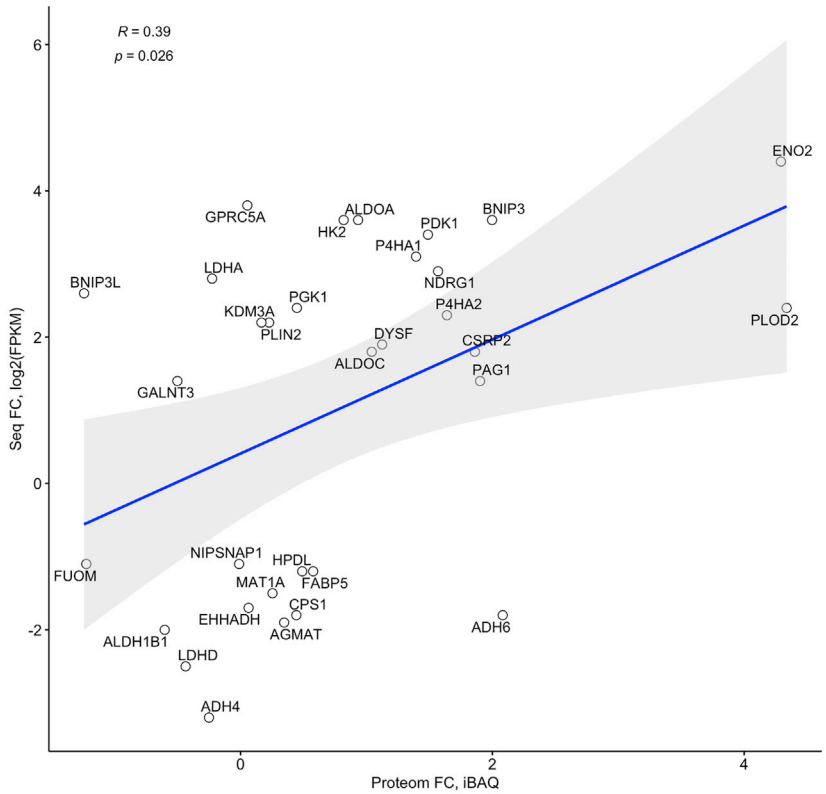


FIGURE 4 | Correlation of fold changes in the RNA sequencing (Seq FC) and proteome (Proteom FC) data in differentiated Caco-2 (dCaco-2) upon cobalt (II) chloride ($CoCl_2$) treatment.

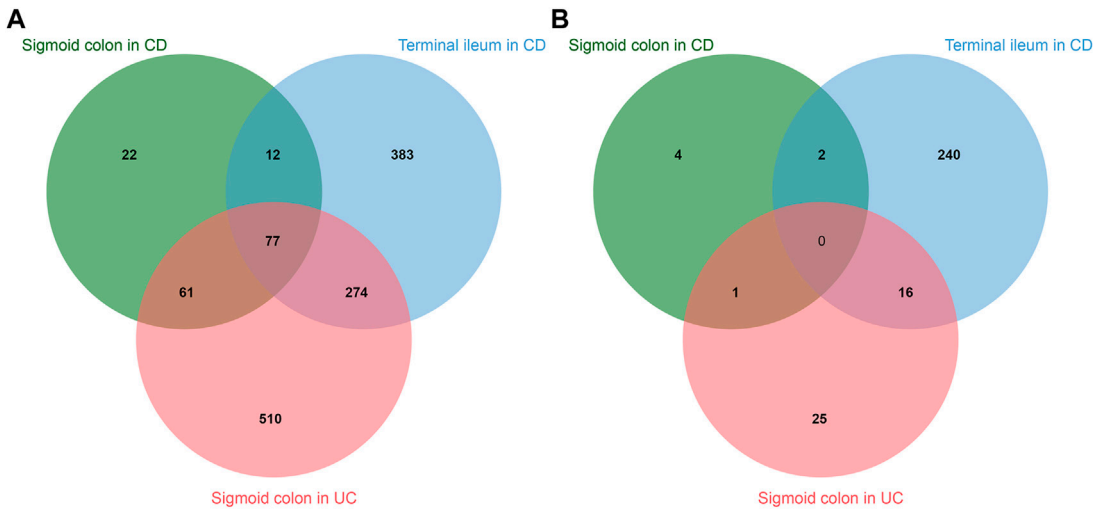


FIGURE 5 | Intersections of differentially expressed genes in sigmoid colon of patients with Chron's disease (CD), terminal ileum of patients with CD, and sigmoid colon of patients with ulcerative colitis (UC). **(A)**—upregulated genes. **(B)**—downregulated genes.

transcription factors (Makarova et al., 2014). This makes it reasonable to search for a specific gene set responding to hypoxia in a particular cell type, as for the intestinal epithelium in our study.

$CoCl_2$ and OD are known HIF inductors, acting through inhibition of HIF prolyl hydroxylases (Knyazev et al., 2021), so their effect on cells is regarded as a chemical model of hypoxia. Gene set enrichment analysis of differentially expressed genes in

Caco-2 intestinal barrier model upon $CoCl_2$ and OD exposure confirmed activation of hypoxia response and also an inflammatory response, proliferation, and apoptosis pathways, which are known effects of HIF activation (Corrado and Fontana, 2020).

Analysis of differentially expressed genes in the Caco-2 model of intestinal hypoxia revealed 115 upregulated and 69 downregulated genes. Enrichment analysis of this gene list confirmed HIF-1 signaling pathway activation and showed involvement of carbon metabolism and glycolysis/gluconeogenesis. HIF-dependent regulation of glycolytic, carbohydrate and fatty acid metabolism is a known process in the intestinal epithelium (Konjar et al., 2021).

Not always activation of a gene at the transcriptional level means activation at the translation level, so proteome data do not always correlate with transcriptomic data (Wang et al., 2019). We confirmed a significant correlation of gene expression at mRNA and protein level in Caco-2 cells exposed to $CoCl_2$. It allows us to use RNA-sequencing data from the hypoxia experiment to search for HIF-dependent genes in samples of intestinal epithelium in IBD.

We searched for co-directed significant expression changes in the intestinal hypoxia model and intestinal epithelium in IBD. Possible HIF-associated gene list from Caco-2 model has seven genes in common with terminal ileum of patients with CD, two genes with sigmoid colon of patients with CD, and eleven genes with sigmoid colon of patients with UC. According to gene set enrichment analysis, the HIF signaling pathway was significantly activated in CD terminal ileum and UC sigmoid colon, which have three activated genes in common: *ITGA5*, *PLAUR*, and *DYSF*. *BHLHE40* was upregulated for sigmoid colon in both UC and CD. Only *PLAUR*, encoding uPAR, was activated in all three clinical specimen groups. Comparison of differentially expressed genes from the KEGG HIF-1 pathway for the Caco-2 model, CD terminal ileum and UC sigmoid colon revealed intersections, with the most prominent changes for *PFKFB3*, *TIMP1*, *ANGPT2*, *HK3*, *SERPINE1*, and *TLR4*.

BHLHE40 encodes Basic Helix-Loop-Helix Family Member E40 that is a stress-inducible transcription factor and is activated by HIF and p53, regulating cell survival and proliferation (Kiss et al., 2020). It is a part of core response to hypoxia gene set (Benita et al., 2009) and regulates inflammatory response in the IBD model (Yu et al., 2018). *DYSF* encodes dysferlin that is a type-II transmembrane protein and is expressed in various tissues, primarily in muscle, controlling membrane repair and vesicle trafficking (Barefield et al., 2021). Its role in IBD is unclear, but elevated levels have been found in the blood of patients with IBD (Ostrowski et al., 2019).

Gene encoding uPAR was upregulated in all IBD samples. The three domains of the uPAR molecule form a cavity for the ligand of this receptor, the urokinase-type plasminogen activator (uPA). In this regard, uPAR is one of the key regulators of pericellular proteolysis, participating in the extracellular matrix remodeling and cell migration regulation (Gorrasi et al., 2020). Extracellular matrix degradation by protease systems is increased in inflamed and damaged tissues, including in IBD, which is an important component of tissue repair control (Genua et al., 2015).

uPAR can interact with vitronectin and integrins, leading to intracellular signal activation and regulating cell proliferation and survival (Smith and Marshall, 2010). An interaction between purified uPAR and $\alpha5\beta1$ integrin has been shown *in vitro* (Wei et al., 2005), with the D3 domain of the uPAR molecule possibly forming the integrin-binding site (Chaurasia et al., 2006; Tang et al., 2008). Co-immunoprecipitation of uPAR and integrin α5β1 has also been detected (Ghiso et al., 1999; Wei et al., 2001). uPAR enhances the binding of integrin α5β1 to fibronectin (Ghiso et al., 1999; Monaghan et al., 2004) and increases cell proliferation through integrin $\alpha5\beta1$ signaling (Aguirre-Ghiso et al., 2001; Monaghan-Benson and McKeown-Longo, 2006). The complex of $\alpha5\beta1$ integrin and uPAR can lead to activation of the epidermal growth factor receptor signaling pathway (Aguirre Ghiso, 2002; Liu et al., 2002). There was a significant increase in the expression of the *ITGA5* gene encoding integrin α5 in Caco-2 cells upon HIF stabilization, in terminal ileum in CD, and in sigmoid colon in UC, which together with the increased expression of uPAR may lead to increased signaling through these molecules.

The role of uPAR in the IBD pathogenesis remains unclear. uPAR is expressed in intestinal crypts in normal state and IBD, although no significant expression difference has been detected (Gibson and Rosella, 1996). The intestinal nerve tissue expresses uPAR in IBD, in contrast to the nerve tissue of the healthy intestine (Laerum et al., 2008). In an animal model of IBD, increased uPAR expression at the protein and mRNA levels was detected in intestinal tissue, and this increase was predominantly due to macrophages in the intestinal wall (Genua et al., 2015). In our experiment, we observed increased *PLAUR* expression in dCaco-2 and uCaco-2 cells as a model of intestinal epithelium upon hypoxia exposure and HIF activation. Howell et al. used magnetic bead sorting for the epithelial cell adhesion molecule to sequence intestinal epithelium in IBD and healthy control (Howell et al., 2018), indicating activation of uPAR expression specifically in the intestinal epithelium.

On the one hand, HIF-activated uPAR expression stimulates extracellular matrix degradation promoting epithelial barrier destruction and cell invasion in carcinomas (Büchler et al., 2009). On the other hand, the uPAR pathway is required for efficient epithelial wound repair by stimulating extracellular matrix remodeling and cell migration (Stewart et al., 2012). It is likely that just as varying degrees of hypoxia can both promote and impair gut barrier function, so varying degrees of uPAR activation can be both damaging and regenerative. Mice with uPAR knockdown were more susceptible to the development of IBD in an experimental model (Genua et al., 2015), which may indicate a protective role of uPAR during HIF activation and inflammation. The total level of uPAR increases in patients with IBD, but the level of membrane-associated uPAR decreases as proteolysis and accumulation of the soluble form of uPAR occurs, which does not perform its functions in the activation of intracellular signaling pathways (Genua et al., 2015) This is consistent with the elevated blood level of the soluble form of uPAR in IBD (Lönnkvist et al., 2011; Kolho et al., 2012). Thus, increased uPAR expression in the intestinal epithelium may be associated not with the destruction of the extracellular matrix and damage to the intestinal wall but rather with damage adaptation, stimulation of migration and proliferation to repair the epithelial

layer through interaction with integrins. Pharmacological activation of the HIF-signaling pathway stimulated intestinal epithelium repair regulating integrin expression and function (Goggins et al., 2021).

HIF-1 activation can upregulate uPAR expression (Büchler et al., 2009; Nishi et al., 2016). Hypoxia also enhances the expression of *ITGA5* (Ju et al., 2017). We identified possible transcription regulators of *ITGA5* and *PLAUR* with the previously developed tool miRGTF-net (Nersisyan et al., 2021b). HIF1A and HIF3A are both transcription factors that can regulate these genes. Analysis of correlation of all detected transcription factors with *ITGA5* and *PLAUR* in the colorectal cancer samples from TCGA revealed six genes with significant correlation. Of these six genes, only *NFATC1* expression has a tendency to be upregulated in uCaco-2 and dCaco-2 upon HIF-pathway activation, suggesting its regulatory role, and *NFATC1* expression is regulated by HIF1A according to miRGTF-net. This suggestion is supported by NFATC1 activation in hypoxic conditions (Shin et al., 2015; Xiao et al., 2020).

5 CONCLUSION

Hypoxia is a common feature in IBD pathogenesis. Chemical hypoxia induction in the Caco-2 intestinal model with cobalt (II) chloride and oxyquinoline derivative revealed potential intestinal HIF-associated gene core. Of that gene core, *ITGA5* and *PLAUR* genes were also upregulated in intestinal epithelial cells in IBD, suggesting activation of tissue regeneration program upon HIF activation. HIF-dependent upregulation of transcription factor NFATC1 is a possible mechanism of *ITGA5* and *PLAUR* activation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material.**

AUTHOR CONTRIBUTIONS

DM conceptualized and reviewed the work. MS worked on analysis and visualization of the bioinformatics data presented. MR carried out laboratory experiments. EK carried out laboratory experiments, worked on curation, analysis and visualization of the data presented, prepared the original draft, edited and reviewed it. All authors contributed to the article and approved the submitted version.

FUNDING

The reported study was funded by RFBR, project number 20-34-70092 (EK, DM, MR) and Laboratory of Molecular Physiology at HSE University (MS).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.791640/full#supplementary-material

REFERENCES

Aguirre Ghiso, J. A. (2002). Inhibition of FAK Signaling Activated by Urokinase Receptor Induces Dormancy in Human Carcinoma Cells *In Vivo*. *Oncogene* 21, 2513–2524. doi:10.1038/sj.onc.1205342

Aguirre-Ghiso, J. A., Liu, D., Mignatti, A., Kovalski, K., and Ossowski, L. (2001). Urokinase Receptor and Fibronectin Regulate the ERKMAPK to p38MAPK Activity Ratios that Determine Carcinoma Cell Proliferation or Dormancy *In Vivo*. *Mol. Biol. Cel* 12, 863–879. doi:10.1091/mbc.12.4.863

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). Jvenn: an Interactive Venn Diagram Viewer. *BMC Bioinformatics* 15, 293. doi:10.1186/1471-2105-15-293

Barefield, D. Y., Sell, J. J., Tahtah, I., Kearns, S. D., McNally, E. M., and Demonbreun, A. R. (2021). Loss of Dysferlin or Myoferlin Results in Differential Defects in Excitation-Contraction Coupling in Mouse Skeletal Muscle. *Sci. Rep.* 11, 15865. doi:10.1038/s41598-021-95378-9

Benita, Y., Kikuchi, H., Smith, A. D., Zhang, M. Q., Chung, D. C., and Xavier, R. J. (2009). An Integrative Genomics Approach Identifies Hypoxia Inducible Factor-1 (HIF-1)-Target Genes that Form the Core Response to Hypoxia. *Nucleic Acids Res.* 37, 4587–4602. doi:10.1093/nar/gkp425

Borg-Bartolo, S. P., Boyapati, R. K., Satsangi, J., and Kalla, R. (2020). Precision Medicine in Inflammatory Bowel Disease: Concept, Progress and Challenges. *F1000Res* 9, 54. doi:10.12688/f1000research.20928.1

Büchler, P., Reber, H. A., Tomlinson, J. S., Hankinson, O., Kallifatidis, G., Friess, H., et al. (2009). Transcriptional Regulation of Urokinase-type Plasminogen Activator Receptor by Hypoxia-Inducible Factor 1 Is Crucial for Invasion of Pancreatic and Liver Cancer. *Neoplasia* 11, 196–IN12. doi:10.1593/neo.08734

Cancer Genome Atlas Network (2012). Comprehensive Molecular Characterization of Human colon and Rectal Cancer. *Nature* 487, 330–337. doi:10.1038/nature11252

Chaurasia, P., Aguirre-Ghiso, J. A., Liang, O. D., Gardsvoll, H., Ploug, M., and Ossowski, L. (2006). A Region in Urokinase Plasminogen Receptor Domain III Controlling a Functional Association with α5β1 Integrin and Tumor Growth. *J. Biol. Chem.* 281, 14852–14863. doi:10.1074/jbc.M512311200

Corrado, C., and Fontana, S. (2020). Hypoxia and HIF Signaling: One Axis with Divergent Effects. *Int. J. Mol. Sci.* 21, 5611. doi:10.3390/ijms21165611

Ding, X., Hu, X., Chen, Y., Xie, J., Ying, M., Wang, Y., et al. (2021). Differentiated Caco-2 Cell Models in Food-Intestine Interaction Study: Current Applications and Future Trends. *Trends Food Sci. Technol.* 107, 455–465. doi:10.1016/j.tifs.2020.11.015

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., et al. (2019). GENCODE Reference Annotation for the Human and Mouse Genomes. *Nucleic Acids Res.* 47, D766–D773. doi:10.1093/nar/gky955

Genua, M., D'Alessio, S., Cibella, J., Gandelli, A., Sala, E., Correale, C., et al. (2015). The Urokinase Plasminogen Activator Receptor (uPAR) Controls Macrophage Phagocytosis in Intestinal Inflammation. *Gut* 64, 589–600. doi:10.1136/gutjnl-2013-305933

Ghiso, J. A. A., Kovalski, K., and Ossowski, L. (1999). Tumor Dormancy Induced by Downregulation of Urokinase Receptor in Human Carcinoma Involves Integrin and MAPK Signaling. *J. Cel Biol.* 147, 89–104. doi:10.1083/jcb.147.1.89

Gibson, P. R., and Rosella, O. (1996). Abnormalities of the Urokinase System in Colonic Crypt Cells from Patients with Ulcerative Colitis. *Inflamm. Bowel Dis.*

2, 105–114. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23282516. doi:10.1097/00054725-199606000-00006

Glover, L. E., Lee, J. S., and Colgan, S. P. (2016). Oxygen Metabolism and Barrier Regulation in the Intestinal Mucosa. *J. Clin. Invest.* 126, 3680–3688. doi:10.1172/JCI84429

Goggins, B. J., Minahan, K., Sherwin, S., Soh, W. S., Pryor, J., Bruce, J., et al. (2021). Pharmacological HIF-1 Stabilization Promotes Intestinal Epithelial Healing through Regulation of α-Integrin Expression and Function. *Am. J. Physiol. Liver Physiol.* 320, G420–G438. doi:10.1152/ajpgi.00192.2020

Gorrasi, A., Petrone, A. M., Li Santi, A., Alfieri, M., Montuori, N., and Ragno, P. (2020). New Pieces in the Puzzle of uPAR Role in Cell Migration Mechanisms. *Cells* 9, 2531. doi:10.3390/cells9122531

Howell, K. J., Kraiczy, J., Nayak, K. M., Gasparetto, M., Ross, A., Lee, C., et al. (2018). DNA Methylation and Transcription Patterns in Intestinal Epithelial Cells from Pediatric Patients with Inflammatory Bowel Diseases Differentiate Disease Subtypes and Associate with Outcome. *Gastroenterology* 154, 585–598. doi:10.1053/j.gastro.2017.10.007

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic Acids Res.* 37, 1–13. doi:10.1093/nar/gkn923

Ivanenkov, Y. A., Vasilevski, S. V., Beloglazkina, E. K., Kukushkin, M. E., Machulkin, A. E., Veselov, M. S., et al. (2015). Design, Synthesis and Biological Evaluation of Novel Potent MDM2/p53 Small-Molecule Inhibitors. *Bioorg. Med. Chem. Lett.* 25, 404–409. doi:10.1016/j.bmcl.2014.09.070

Ju, J. A., Godet, I., Ye, I. C., Byun, J., Jayatilaka, H., Lee, S. J., et al. (2017). Hypoxia Selectively Enhances Integrin α5β1 Receptor Expression in Breast Cancer to Promote Metastasis. *Mol. Cancer Res.* 15, 723–734. doi:10.1158/1541-7786.MCR-16-0338

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: Integrating Viruses and Cellular Organisms. *Nucleic Acids Res.* 49, D545–D551. doi:10.1093/nar/gkaa970

Kerber, E. L., Padberg, C., Koll, N., Schuetzhold, V., Fandrey, J., and Winning, S. (2020). The Importance of Hypoxia-Inducible Factors (HIF-1 and HIF-2) for the Pathophysiology of Inflammatory Bowel Disease. *Int. J. Mol. Sci.* 21, 8551. doi:10.3390/ijms21228551

Kim, Y.-E., Lee, M., Gu, H., Kim, J., Jeong, S., Yeo, S., et al. (2018). Hypoxia-inducible Factor-1 (HIF-1) Activation in Myeloid Cells Accelerates DSS-Induced Colitis Progression in Mice. *Dis. Model. Mech.* 11, dmm033241. doi:10.1242/dmm.033241

Kim, Y.-I., Yi, E.-J., Kim, Y.-D., Lee, A. R., Chung, J., Ha, H. C., et al. (2021). Local Stabilization of Hypoxia-Inducible Factor-1α Controls Intestinal Inflammation via Enhanced Gut Barrier Function and Immune Regulation. *Front. Immunol.* 11, 609689. doi:10.3389/fimmu.2020.609689

Kiss, Z., Mudryj, M., and Ghosh, P. M. (2020). Non-Circadian Aspects of BHLHE40 Cellular Function in Cancer. *Genes Cancer* 11, 1–19. doi:10.18632/genesandcancer.201

Knyazev, E., and Paul, S. (2021). Levels of miR-374 Increase in BeWo B30 Cells Exposed to Hypoxia. *Bull. Russ. State. Med. Univ.* 11, 17. doi:10.24075/brsmu.2021.021

Knyazev, E. N., Petrov, V. A., Gazizov, I. N., Gerasimenko, T. N., Tsypina, I. M., Tonevitsky, A. G., et al. (2019a). Oxyquinoline-Dependent Changes in Claudin-Encoding Genes Contribute to Impairment of the Barrier Function of the Trophoblast Monolayer. *Bull. Exp. Biol. Med.* 166, 369–372. doi:10.1007/s10517-019-04352-z

Knyazev, E. N., Zakharova, G. S., Astakhova, L. A., Tsypina, I. M., Tonevitsky, A. G., and Sukhikh, G. T. (2019b). Metabolic Reprogramming of Trophoblast Cells in Response to Hypoxia. *Bull. Exp. Biol. Med.* 166, 321–325. doi:10.1007/s10517-019-04342-1

Knyazev, E. N., Paul, S. Y., and Tonevitsky, A. G. (2021). Chemical Induction of Trophoblast Hypoxia by Cobalt Chloride Leads to Increased Expression of DDIT3. *Dokl. Biochem. Biophys.* 499, 251–256. doi:10.1134/S1607672921040104

Kolho, K.-L., Valtonen, E., Rintamäki, H., and Savilahti, E. (2012). Soluble Urokinase Plasminogen Activator Receptor suPAR as a Marker for Inflammation in Pediatric Inflammatory Bowel Disease. *Scand. J. Gastroenterol.* 47, 951–955. doi:10.3109/00365521.2012.699549

Konjar, Š., Pavšič, M., and Veldhoen, M. (2021). Regulation of Oxygen Homeostasis at the Intestinal Epithelial Barrier Site. *Int. J. Mol. Sci.* 22, 9170. doi:10.3390/ijms22179170

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., and Sergushichev, A. (2016). Fast Gene Set Enrichment Analysis. *bioRxiv*. doi:10.1101/060012

Laerum, O. D., Illemann, M., Skarstein, A., Helgeland, L., Øvrebø, K., Danø, K., et al. (2008). Crohn's Disease but Not Chronic Ulcerative Colitis Induces the Expression of PAI-1 in Enteric Neurons. *Am. J. Gastroenterol.* 103, 2350–2358. doi:10.1111/j.1572-0241.2008.01930.x

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cel Syst.* 1, 417–425. doi:10.1016/j.cels.2015.12.004

Liu, D., Ghiso, J. A. A., Estrada, Y., and Ossowski, L. (2002). EGFR Is a Transducer of the Urokinase Receptor Initiated Signal that Is Required for *In Vivo* Growth of a Human Carcinoma. *Cancer Cell* 1, 445–457. doi:10.1016/s1535-6108(02)00072-7

Lönnkvist, M. H., Theodorsson, E., Holst, M., Ljung, T., and Hellström, P. M. (2011). Blood Chemistry Markers for Evaluation of Inflammatory Activity in Crohn's Disease during Infliximab Therapy. *Scand. J. Gastroenterol.* 46, 420–427. doi:10.3109/00365521.2010.539253

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Makarova, J. A., Maltseva, D. V., Galatenko, V. V., Abbasi, A., Maximenko, D. G., Grigoriev, A. I., et al. (2014). Exercise Immunology Meets MiRNAs. *Exerc. Immunol. Rev.* 20, 135–164. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24974725.

Maltseva, D., Poloznikov, A., and Artyushenko, V. (2020). Selective Changes in Expression of Integrin α-Subunits in the Intestinal Epithelial Caco-2 Cells under Conditions of Hypoxia and Microcirculation. *Bull. Russ. State. Med. Univ.* 23, 30. doi:10.24075/brsmu.2020.078

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet j.* 17, 10. doi:10.14806/ej.17.1.200

Monaghan, E., Gueorguiev, V., Wilkins-Port, C., and McKeown-Longo, P. J. (2004). The Receptor for Urokinase-type Plasminogen Activator Regulates Fibronectin Matrix Assembly in Human Skin Fibroblasts. *J. Biol. Chem.* 279, 1400–1407. doi:10.1074/jbc.M310374200

Monaghan-Benson, E., and McKeown-Longo, P. J. (2006). Urokinase-Type Plasminogen Activator Receptor Regulates a Novel Pathway of Fibronectin Matrix Assembly Requiring Src-dependent Transactivation of Epidermal Growth Factor Receptor. *J. Biol. Chem.* 281, 9450–9459. doi:10.1074/jbc.M501901200

Muñoz-Sánchez, J., and Chánez-Cárdenas, M. E. (2019). The Use of Cobalt Chloride as a Chemical Hypoxia Model. *J. Appl. Toxicol.* 39, 556–570. doi:10.1002/jat.3749

Mutsenko, V. V., Bazhenov, V. V., Rogulska, O., Tarusin, D. N., Schütz, K., Brüggemeier, S., et al. (2017). 3D Chitinous Scaffolds Derived from Cultivated marine Demosponge Aplysina Aerophoba for Tissue Engineering Approaches Based on Human Mesenchymal Stromal Cells. *Int. J. Biol. Macromolecules* 104, 1966–1974. doi:10.1016/j.ijbiomac.2017.03.116

Nersisyan, S., Galatenko, A., Chekova, M., and Tonevitsky, A. (2021a). Hypoxia-Induced miR-148a Downregulation Contributes to Poor Survival in Colorectal Cancer. *Front. Genet.* 12, 662468. doi:10.3389/fgene.2021.662468

Nersisyan, S., Galatenko, A., Galatenko, V., Shkurnikov, M., and Tonevitsky, A. (2021b). miRGTF-Net: Integrative miRNA-Gene-TF Network Analysis Reveals Key Drivers of Breast Cancer Recurrence. *PLoS One* 16, e0249424. doi:10.1371/journal.pone.0249424

Nikulin, S. V., Knyazev, E. N., Gerasimenko, T. N., Shilin, S. A., Gazizov, I. N., Zakharova, G. S., et al. (2018a). Non-Invasive Evaluation of Extracellular Matrix Formation in the Intestinal Epithelium. *Bull. Exp. Biol. Med.* 166, 35–38. doi:10.1007/s10517-018-4283-7

Nikulin, S. V., Knyazev, E. N., Poloznikov, A. A., Shilin, S. A., Gazizov, I. N., Zakharova, G. S., et al. (2018b). Expression of SLC30A10 and SLC23A3 Transporter mRNAs in Caco-2 Cells Correlates with an Increase in the Area of the Apical Membrane. *Mol. Biol.* 52, 577–582. doi:10.1134/S0026893318040131

Knyazev, E. N., Maltseva, D. V., Sakharov, D. A., Gerasimenko, T. N., and Gerasimenko, T. N. (2019). Use of Impedance Spectroscopy to Assess the Effect of Laminins on the *In Vitro* Differentiation of Intestinal Cells. *Biiotehnologiâ (Mosk.)* 35, 102–107. doi:10.21519/0234-2758-2019-35-6-102-107

Nishi, H., Sasaki, T., Nagamitsu, Y., Terauchi, F., Nagai, T., Nagao, T., et al. (2016). Hypoxia Inducible Factor-1 Mediates Upregulation of Urokinase-type Plasminogen Activator Receptor Gene Transcription during Hypoxia in Cervical Cancer Cells. *Oncol. Rep.* 35, 992–998. doi:10.3892/or.2015.4449

Ostrowski, J., Goryca, K., Goryca, K., Lazowska, I., Rogowska, A., Paziewska, A., et al. (2019). Common Functional Alterations Identified in Blood Transcriptome of Autoimmune Cholestatic Liver and Inflammatory Bowel Diseases. *Sci. Rep.* 9, 7190. doi:10.1038/s41598-019-43699-1

Pral, L. P., Fachi, J. L., Corrêa, R. O., Colonna, M., and Vinolo, M. A. R. (2021). Hypoxia and HIF-1 as Key Regulators of Gut Microbiota and Host Interactions. *Trends Immunol.* 42, 604–621. doi:10.1016/j.it.2021.05.004

Ramakrishnan, S. K., and Shah, Y. M. (2016). Role of Intestinal HIF-2α in Health and Disease. *Annu. Rev. Physiol.* 78, 301–325. doi:10.1146/annurev-physiol-021115-105202

Rangel-Huerta, E., and Maldonado, E. (2017). Transit-Amplifying Cells in the Fast Lane from Stem Cells towards Differentiation. *Stem Cell Int.* 2017, 7602951. doi:10.1155/2017/7602951

Sakharov, D., Maltseva, D., Knyazev, E., Nikulin, S., Poloznikov, A., Shilin, S., et al. (2019). Towards Embedding Caco-2 Model of Gut Interface in a Microfluidic Device to Enable Multi-Organ Models for Systems Biology. *BMC Syst. Biol.* 13, 19. doi:10.1186/s12918-019-0686-y

Samanta, D., and Semenza, G. L. (2017). Maintenance of Redox Homeostasis by Hypoxia-Inducible Factors. *Redox Biol.* 13, 331–335. doi:10.1016/j.redox.2017.05.022

Samatov, T. R., Senyavina, N. V., Galatenko, V. V., Trushkin, E. V., Tonevitskaya, S. A., Alexandrov, D. E., et al. (2016). Tumour-Like Druggable Gene Expression Pattern of CaCo2 Cells in Microfluidic Chip. *Biochip J.* 10, 215–220. doi:10.1007/s13206-016-0308-3

Samatov, T. R., Galatenko, V. V., Senyavina, N. V., Galatenko, A. V., Shkurnikov, M. Y., Tonevitskaya, S. A., et al. (2017). miRNA-Mediated Expression Switch of Cell Adhesion Genes Driven by Microcirculation in Chip. *Biochip J.* 11, 262–269. doi:10.1007/s13206-017-1305-x

Schoultz, I., and Keita, Å. V. (2020). The Intestinal Barrier and Current Techniques for the Assessment of Gut Permeability. *Cells* 9, 1909. doi:10.3390/cells9081909

Shin, J., Nunomiya, A., Kitajima, Y., Dan, T., Miyata, T., and Nagatomi, R. (2015). Prolyl Hydroxylase Domain 2 Deficiency Promotes Skeletal Muscle Fiber-Type Transition via a calcineurin/NFATc1-Dependent Pathway. *Skeletal Muscle* 6, 5. doi:10.1186/s13395-016-0079-5

Singhal, R., and Shah, Y. M. (2020). Oxygen Battle in the Gut: Hypoxia and Hypoxia-Inducible Factors in Metabolic and Inflammatory Responses in the Intestine. *J. Biol. Chem.* 295, 10493–10505. doi:10.1074/jbc.REV120.011188

Smirnov, I., Carletti, E., Kurkova, I., Nachon, F., Nicolet, Y., Mitkevich, V. A., et al. (2011). Reactibodies Generated by Kinetic Selection Couple Chemical Reactivity with Favorable Protein Dynamics. *Proc. Natl. Acad. Sci.* 108, 15954–15959. doi:10.1073/pnas.1108460108

Smith, H. W., and Marshall, C. J. (2010). Regulation of Cell Signalling by uPAR. *Nat. Rev. Mol. Cel Biol.* 11, 23–36. doi:10.1038/nrm2821

Solanki, S., Devenport, S. N., Ramakrishnan, S. K., and Shah, Y. M. (2019). Temporal Induction of Intestinal Epithelial Hypoxia-Inducible Factor-2α Is Sufficient to Drive Colitis. *Am. J. Physiology-Gastrointestinal Liver Physiol.* 317, G98–G107. doi:10.1152/ajpgi.00081.2019

Stewart, C. E., Nijmeh, H. S., Brightling, C. E., and Sayers, I. (2012). uPAR Regulates Bronchial Epithelial Repair *In Vitro* and Is Elevated in Asthmatic Epithelium. *Thorax* 67, 477–487. doi:10.1136/thoraxjnl-2011-200508

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. doi:10.1073/pnas.0506580102

Tang, C.-H., Hill, M. L., Brumwell, A. N., Chapman, H. A., and Wei, Y. (2008). Signaling through Urokinase and Urokinase Receptor in Lung Cancer Cells Requires Interactions with β1 Integrins. *J. Cel Sci.* 121, 3747–3756. doi:10.1242/jcs.029769

Taylor, C. T. (2018). Hypoxia in the Gut. *Cell Mol. Gastroenterol. Hepatol.* 5, 61–62. doi:10.1016/j.jcmgh.2017.09.005

Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D. P., et al. (2019). A Deep Proteome and Transcriptome Abundance Atlas of 29 Healthy Human Tissues. *Mol. Syst. Biol.* 15, e8503. doi:10.15252/msb.20188503

Wei, Y., Eble, J. A., Wang, Z., Kreidberg, J. A., and Chapman, H. A. (2001). Urokinase Receptors Promote β1 Integrin Function through Interactions with Integrin α3β1. *Mol. Biol. Cell* 12, 2975–2986. doi:10.1091/mbc.12.10.2975

Wei, Y., Czekay, R.-P., Robillard, L., Kugler, M. C., Zhang, F., Kim, K. K., et al. (2005). Regulation of α5β1 Integrin Conformation and Function by Urokinase Receptor Binding. *J. Cel Biol.* 168, 501–511. doi:10.1083/jcb.200404112

Xiao, C., Bai, G., Du, Y., Jiang, H., and Yu, X. (2020). Association of High HIF-1α Levels in Serous Periodontitis with External Root Resorption by the NFATc1 Pathway. *J. Mol. Hist.* 51, 649–658. doi:10.1007/s10735-020-09911-7

Xue, X., Ramakrishnan, S., Anderson, E., Taylor, M., Zimmermann, E. M., Spence, J. R., et al. (2013). Endothelial PAS Domain Protein 1 Activates the Inflammatory Response in the Intestinal Epithelium to Promote Colitis in Mice. *Gastroenterology* 145, 831–841. doi:10.1053/j.gastro.2013.07.010

Yu, F., Sharma, S., Jankovic, D., Gurram, R. K., Su, P., Hu, G., et al. (2018). The Transcription Factor Bhlhe40 Is a Switch of Inflammatory versus Antiinflammatory Th1 Cell Fate Determination. *J. Exp. Med.* 215, 1813–1821. doi:10.1084/jem.20170155

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 12 November 2021
doi: 10.3389/fgene.2021.741112

A Novel Risk-Score Model With Eight MiRNA Signatures for Overall Survival of Patients With Lung Adenocarcinoma

Jun Wu[1†], *Yuqing Lou*[2†], *Yi-Min Ma*[1], *Jun Xu*[3] *and Tieliu Shi*[1,4*]

[1]Center for Bioinformatics and Computational Biology, And the Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China, [2]Department of Pulmonary Medicine, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [3]Department of Emergency Medicine, The First Hospital of Anhui Medical University, Hefei, China, [4]Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University and Capital Medical University, Beijing, China

Lung adenocarcinoma (LUAD) is the most common subtype of lung cancer with heterogeneous outcomes and diverse therapeutic responses. To classify patients into different groups and facilitate the suitable therapeutic strategy, we first selected eight microRNA (miRNA) signatures in The Cancer Genome Atlas (TCGA)-LUAD cohort based on multi-strategy combination, including differential expression analysis, regulatory relationship, univariate survival analysis, importance clustering, and multivariate combinations analysis. Using the eight miRNA signatures, we further built novel risk scores based on the predefined cutoff and beta coefficients and divided the patients into high-risk and low-risk groups with significantly different overall survival time (p-value < 2 e−16). The risk-score model was confirmed with an independent dataset (p-value = 4.71 e−4). We also observed that the risk scores of early-stage patients were significantly lower than those of late-stage patients. Moreover, our model can also provide new insights into the current clinical staging system and can be regarded as an alternative system for patient stratification. This model unified the variable value as the beta coefficient facilitating the integration of biomarkers obtained from different omics data.

Keywords: lung adenocarcinoma, microRNA signature, risk-score model, overall survival time, treatment response

OPEN ACCESS

Edited by:
Wen-Lian Chen,
Shanghai University of Traditional Chinese Medicine, China

Reviewed by:
Fan Yang,
Jiangxi Science and Technology Normal University, China
Rongzhong Huang,
Second Affiliated Hospital of Chongqing Medical University, China

***Correspondence:**
Tieliu Shi
tieliushi@yahoo.com

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *14 July 2021*
Accepted: *08 October 2021*
Published: *12 November 2021*

Citation:
Wu J, Lou Y, Ma Y-M Xu J and Shi T (2021) A Novel Risk-Score Model With Eight MiRNA Signatures for Overall Survival of Patients With Lung Adenocarcinoma. Front. Genet. 12:741112. doi: 10.3389/fgene.2021.741112

INTRODUCTION

Lung cancer, which is one of the most common and severe types of cancer, remains the leading cause of cancer incidence and mortality worldwide in both males and females (Siegel et al., 2019). Lung adenocarcinoma (LUAD) is the most prevalent histological subtype of lung cancer, with an increasing incidence over the past few decades (Ferlay et al., 2010). The traditional clinical staging system for LUAD, which is based on anatomical information, appears to be inadequate for prognosis evaluation or treatment choices now due to the heterogeneity among patients.

With the rapid advance of molecular biology, many diagnostic and prognostic biomarkers have been identified for various cancers (Wang et al., 2017a; Wang et al., 2017b; Cheng et al., 2019; Huang et al., 2020a; Sheng et al., 2020). With the use of these biomarkers, the traditional tumor classes can be further divided into new subtypes, which may benefit from different therapeutic strategies (Li et al., 2019; Sherafatian and Arjmand, 2019; Lathwal et al., 2020). Besides that, most targeted agents (e.g.,

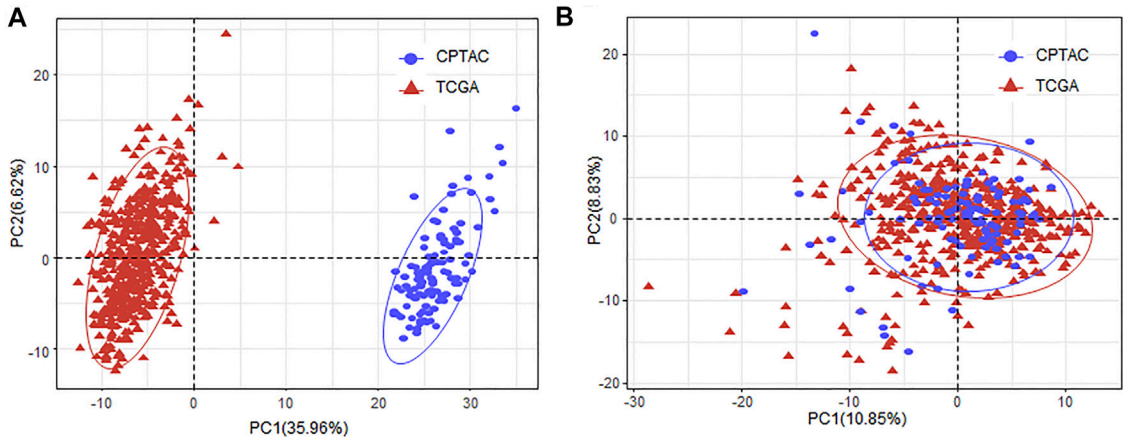


FIGURE 1 | Removing batch effect of the miRNA expression between TCGA and CPTAC datasets. **(A)** PCA plot of the samples obtained from TCGA and CPTAC database with the miRNA expression before batch effect removal. **(B)** PCA plot of the samples obtained from TCGA and CPTAC database with the miRNA expression after batch effect removal. MiRNA, microRNA; TCGA, The Cancer Genome Atlas; CPTAC, Clinical Proteomic Tumor Analysis Consortium; PCA, principal component analysis.

cetuximab, gefitinib, and tamoxifen) are effectively only if their respective targets are mutated or differentially expressed (Sun et al., 2017; Yang et al., 2020).

MicroRNAs (miRNAs) are small non-protein-coding RNAs, which can negatively regulate gene expression by binding to their selective messenger RNAs (mRNAs), thereby influencing various biological progresses, such as cellular differentiation, cell-cycle control, and apoptosis (Bentwich, 2005; Cheng et al., 2005; Novello et al., 2013). MiRNAs are reported to be differentially expressed in various human cancers and act as both tumor suppressors and oncogenes (Volinia et al., 2006; Cui et al., 2020). For some certain types of cancer, the miRNAs are proved to be more effective in cancer classification than mRNAs (Miska, 2007), and the miRNAs are also used as signatures for prognosis prediction. Yu et al. identified five miRNAs significantly associated with patient relapse and survival based on 117 non-small cell lung cancer (NSCLC) patients (Yu et al., 2008). Li et al. also identified eight miRNAs as signatures for survival prediction in LUAD (Li et al., 2014). Similarly, Hess et al. provided a five-miRNA signature, which is a strong and independent prognostic factor for disease recurrence and survival of patients with HPV-negative head and neck squamous cell carcinoma (HNSCC) (Hess et al., 2019). All these results showed that miRNAs are powerful potential signatures for prognosis prediction. However, there were very few overlaps between these miRNA signatures identified by different groups. Moreover, most studies just focused on the miRNA or mRNA expression level independently and ignored the negatively regulative relationship between miRNAs and mRNAs.

In this study, based on the miRNA expression, gene expression profiles and clinical information of 516 LUAD samples from The Cancer Genome Atlas (TCGA) (The Cancer Genome Atlas Research Network, 2014), we built the miRNA–gene negative regulation pairs to ensure that the candidate miRNAs influence biological progress of these samples. Then, we screened eight miRNA signatures through differential expression analysis, regulatory relationship filtering, univariate survival analysis, importance clustering, and multivariate combination selection. Based on the eight miRNA signatures, we built a risk-score model to group the patients as high-risk and low-risk. The model performance was further proved using an independent dataset. We demonstrated that the model can also be used for stratification of patients in the same tumor stage.

RESULTS

Data Collection

The gene expression, miRNA expression, and clinical data of TCGA-LUAD were download from UCSC Xena (http://xena.ucsc.edu) (Goldman et al., 2017). Besides that, we also downloaded the miRNA expression and related clinical data of LUAD from the Clinical Proteomic Tumor Analysis Consortium (CPTAC)-3 database (Edwards et al., 2015) using the R/Biconductor package "TCGAbiolinks" as the independent validation data (Colaprico et al., 2016; Mounir et al., 2019). Only the primary solid tumor (TP) and solid tissue normal (NT) samples were selected. Patients with less than 30 days of overall survival (OS) were excluded to avoid the possible unrelated causes of death. The details of the samples are shown in **Table 1**.

As the miRNA expression was obtained from different databases, we applied ComBat (Leek et al., 2012) to remove the batch effect (**Figures 1A,B**).

Differential Gene Expression Analysis

The count data of gene expression were used to perform the differential expression analysis. The genes with adjusted *p*-value of less than 1 e−3 and absolute log2 fold change ≥1 were regarded as significantly differentially expressed. As a result, a total of 4,522 (64.11%) upregulated and 2,531 (35.89%) downregulated genes

TABLE 1 | Number of samples obtained from different databases.

TCGA-LUAD				CPTAC-LUAD	
Gene expression		MiRNA expression		MiRNA expression	
TP	NT	TP	NT	TP	NT
510	58	510	45	111	102

Note. OS, overall survival; TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma; CPTAC, Clinical Proteomic Tumor Analysis Consortium; miRNA, microRNA; TP, primary solid tumor; NT, solid tissue normal.

(**Figure 2A**). The Gene Ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis results showed that these differentially expressed genes (DEGs) were enriched in 842 biological processes (BPs), 161 molecular functions (MFs), 137 cellular components (CCs), and 44 KEGG pathways (**Figure 2B**; **Supplement Table S1**).

MicroRNA Signature Identification Based on Multi-Strategy

Using the negative regulation criterion and the information retrieved from three verified miRNA-target databases, we obtained 2,284 miRNA–gene pairs consisting of 228 miRNAs and 1,199 target genes. To examine the function term and effects of these miRNA regulators, we performed GO term and pathway enrichment analysis for these 1,199 target genes. The results showed that there were 924 genes functionally enriched in 700 BPs, 30 MFs, and 53 CCs with adjusted *p*-value of less than 0.05 (**Supplement Table S2**). Additionally, there were 163 genes enriched in 16 KEGG pathways, such as cell cycle, cellular senescence, and p53 signaling pathway (**Supplement Table S2**). By limiting the target genes as these functional enriched genes, we simplified the miRNA–gene regulation network consisting of 221 miRNA and 924 genes (**Figure 3A**).

We next performed the univariate survival analysis using the Cox proportional-hazards model with the 161 miRNA regulators. The results showed that 20 miRNAs of LUAD patients can be divided into two groups with significantly different OS (adjusted *p*-value of less than 0.05, **Supplement Table S3**). To further ensure the robustness of these miRNAs, we repeatedly performed survival analysis using randomForestSRC 5,000 times and measured the importance of the 21 miRNAs accordingly. With the variable importance rank matrix (see *Methods*), we clustered the 21 miRNAs into three groups using hierarchical cluster analysis (**Figure 3B**), and 13 miRNAs that ranked top in most of the repeats were selected for the downstream analysis.

To further select the optimal combination of the miRNA signatures, we performed multivariate survival analysis by adding the 13 miRNAs into the Cox regression model using greedy strategy (**Figure 3C**). By doing so, we observed that when the number of the miRNA signatures reached eight, the performance was no longer improved. Thus, we selected eight miRNAs (hsa-mir-1293, hsa-mir-4734, hsa-mir-6132, hsa-mir-4487, hsa-mir-4794, hsa-mir-4517, hsa-mir-7705, and hsa-mir-4784) as the miRNA signatures to build the risk-score prediction model.

For each of the miRNA signatures, we divided LUAD patients into two groups according to the miRNA expression with different thresholds and evaluated the discrimination validity

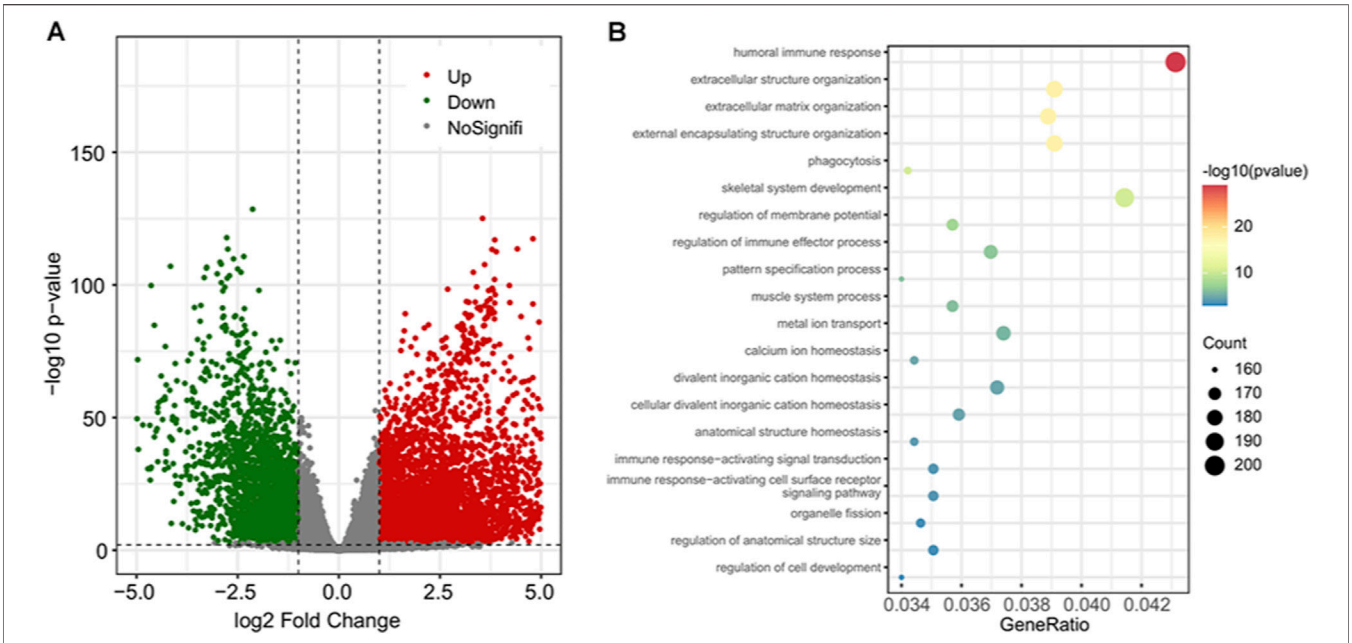


FIGURE 2 | The differential gene expression analysis results and enriched functional terms. **(A)** The volcano plot of the DEGs; 4,522 upregulated genes are in red and 2,531 downregulated genes are in green. **(B)** Bubble plot of the top 20 enriched biological processes. DEGs, differentially expressed genes.

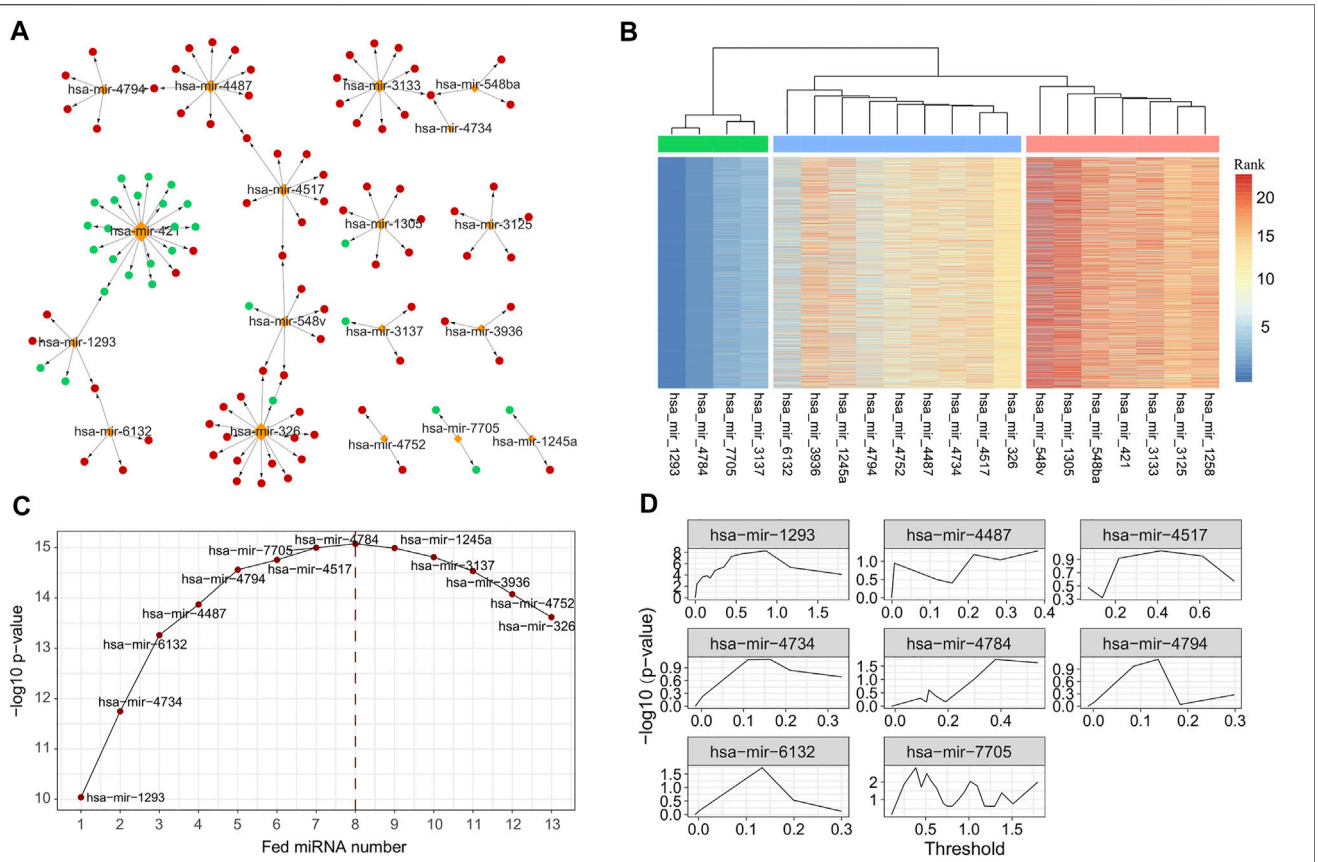


FIGURE 3 | MiRNAs selected with different strategy. **(A)** The subgraph of miRNA–gene regulatory network, consisting of 26 miRNAs selected by univariate survival analysis. **(B)** Heatmap of importance rank obtained with repeatedly performed survival analysis using randomForestSRC 5,000 times. The 26 miRNAs were further clustered into three groups, and 12 miRNAs were regarded as core or important miRNAs. **(C)** Prognostic ability [measured with −log10 (*p*-value)] of miRNA combination generated by feeding the selected 12 core or important miRNAs successively. **(D)** Optimal thresholds selected for the final eight miRNA signatures. MiRNA, microRNA.

using log-rank test and Kaplan–Meier test (**Figure 3D**). The optimal threshold and the β coefficients for each miRNA signature were saved for the model building (see *Methods*).

Performance Evaluation for the Risk-Score Model

Using the risk-score model, we estimated the risk score for each LUAD patient and divided the LUAD cohort into high-risk and low-risk groups by defining the cutoff as the median risk score (cutoff = 2.9). The Kaplan–Meier survival analysis results showed that the OS time was significantly different between the patients in these two groups (*p*-value = 1.43 e−18, **Figure 4A**). We also evaluated the performance with the independent validation dataset (CPTAC-LUAD). The risk score of the patient in the CPTAC-LUAD dataset were estimated, and then the CPTAC-LUAD patients were divided into high-risk and low-risk groups with the cutoff determined by TCGA-LUAD dataset. The Kaplan–Meier survival analysis results showed that the OS time was significantly different between the CPTAC-LUAD patients in these two groups (*p*-value = 4.71 e−4, **Figure 4B**).

To further assess the prognostic power of proposed method, time-dependent receiver operating characteristic (ROC) curves were used to compare the specificity and sensitivity for the predicted results of TCGA-LUAD cohort (1 year, 0.716; 3 years, 0.685; 5 years, 0.657; **Figure 4C**) and CPTAC-LUAD cohort (1 year, 0.693; 3 years, 0.657; **Figure 4D**). The ROC curves and area under the ROC curve (AUC) showed high consistency of this risk-score model.

The Prognostic Ability of the Risk-Score Model Within Different Clinical Groups

To further validate the prognostic ability of the risk-score model, we test the enrichment of low- and high-risk patients in the groups divided by different clinical indicators, such as age, gender, and clinical stages (Stages I–IV). We found that there was no significant difference of the risk score between the male and female patients (*p*-value = 0.133), and the risk score also did not show significant correlation with the patient age (R = −0.079, *p*-value = 0.1, **Figure 5A**). For the clinical stages, we found that the risk score of patients in Stage II and Stage III were significantly higher than that of patients in Stage I (Stage II: *p*-value = 1.2 e−5,

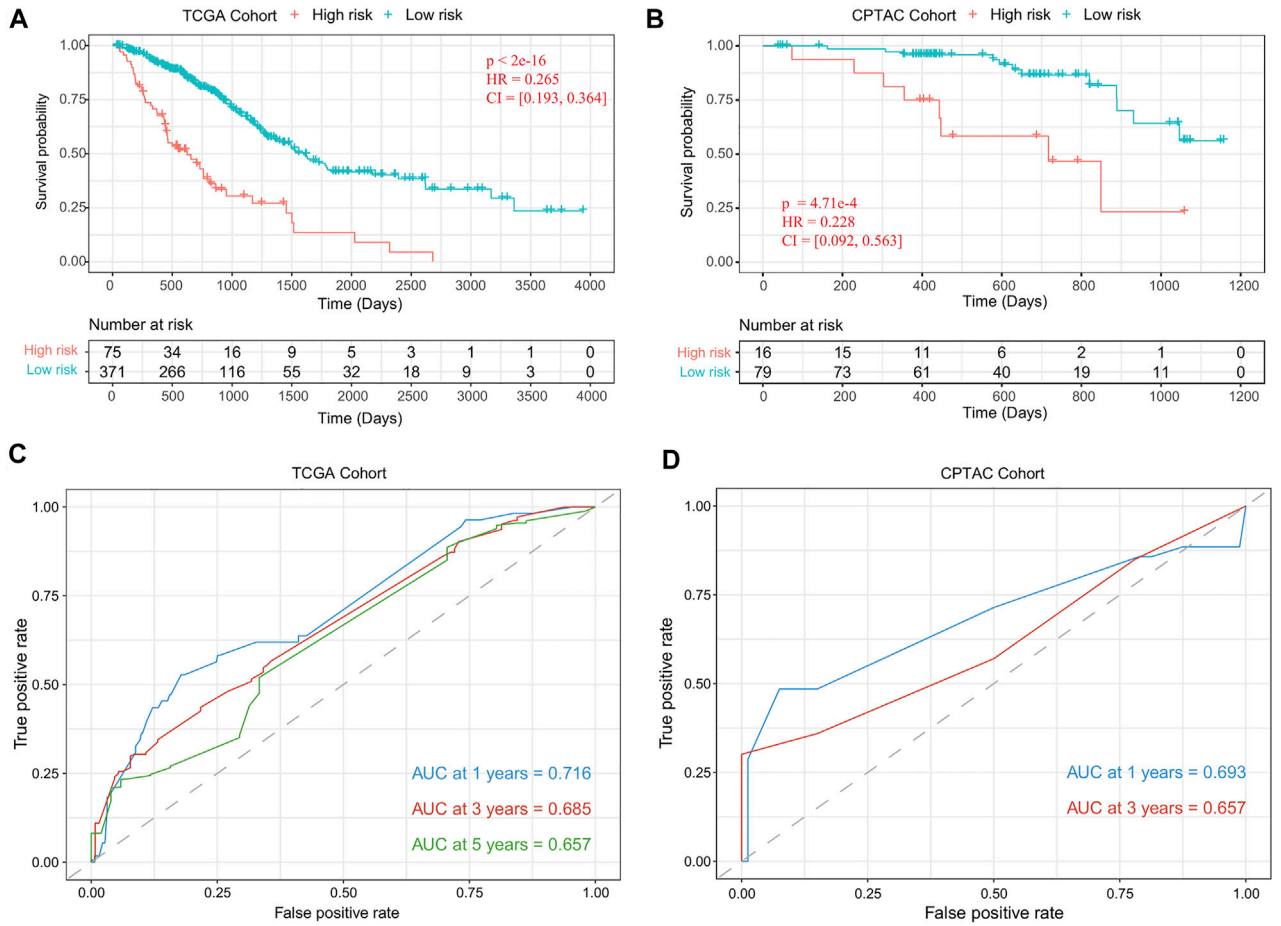


FIGURE 4 | Performance evaluation of the risk-score model. **(A)** Kaplan–Meier plots of OS in TCGA-LUAD cohort when the risk-score cutoff was set as the median value (cutoff = 2.9). **(B)** Kaplan–Meier plots of OS in CPTAC-LUAD cohort when the risk-score cutoff set as 2.9. **(C)** ROC curves of risk-score model for TCGA-LUAD cohort. **(D)** ROC curves of risk-score model for the CPTAC-LUAD cohort. OS, overall survival; TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma; CPTAC, Clinical Proteomic Tumor Analysis Consortium; ROC, receiver operating characteristic.

Stage III: *p*-value = 4.3 e–4, **Figure 5B**). The low-risk patients were significantly enriched in early stage (Wilcoxon rank sum test *p*-value < 2.2 e–16). The clinical staging system is the most acknowledged clinicopathological factor for prognostication and therapy determination of LUAD, which are limited because the prognoses within the same clinical stage vary widely (Mlecnik et al., 2011). To further investigate the potentiality of the risk-score model, we tested the difference of OS between the low- and high-risk patients within the same clinical stage. The results showed that, for Stage I, Stage II, and Stage III, OS time was significantly shorter in the high-risk cohort compared with the low-risk cohort (Stage I, *p*-value = 3.12 e–8; Stage II, *p*-value = 0.05; Stage III, *p*-value = 5.23 e–5; **Figures 5C–E**).

Treatment Response for the Groups Divided by the Risk-Score Model

To further evaluate the clinical benefit of the risk-score model, we extracted the treatment information for the LUAD patients, and 155 patients received different types treatment and 297 patients without any treatment information. Patients who received more than two types of therapy (e.g., patients received both chemotherapy and immunotherapy) were excluded for the follow-up analysis. As the patients who received chemotherapy were enriched in Stage II–Stage IV (Fisher's exact test *p*-value = 2.87 e–24), we test the effectiveness of the chemotherapy on the patients in Stage II–Stage IV. The results showed that chemotherapy can improve prognosis to some extent (*p*-value = 0.09, **Figure 6A**).

We also observed that, in all the patients who received chemotherapy, the patients regarded as low-risk also benefited more from the chemotherapy than the high-risk chemotherapy (*p*-value = 1.5 e–4, **Figure 6B**). In chemotherapy drugs specifically, we also observed that carboplatin can significantly prolong the OS of low-risk patients (*p*-value = 0.02, **Figure 5C**), but it has no benefit in the high-risk patients (*p*-value = 0.94, **Figure 5D**).

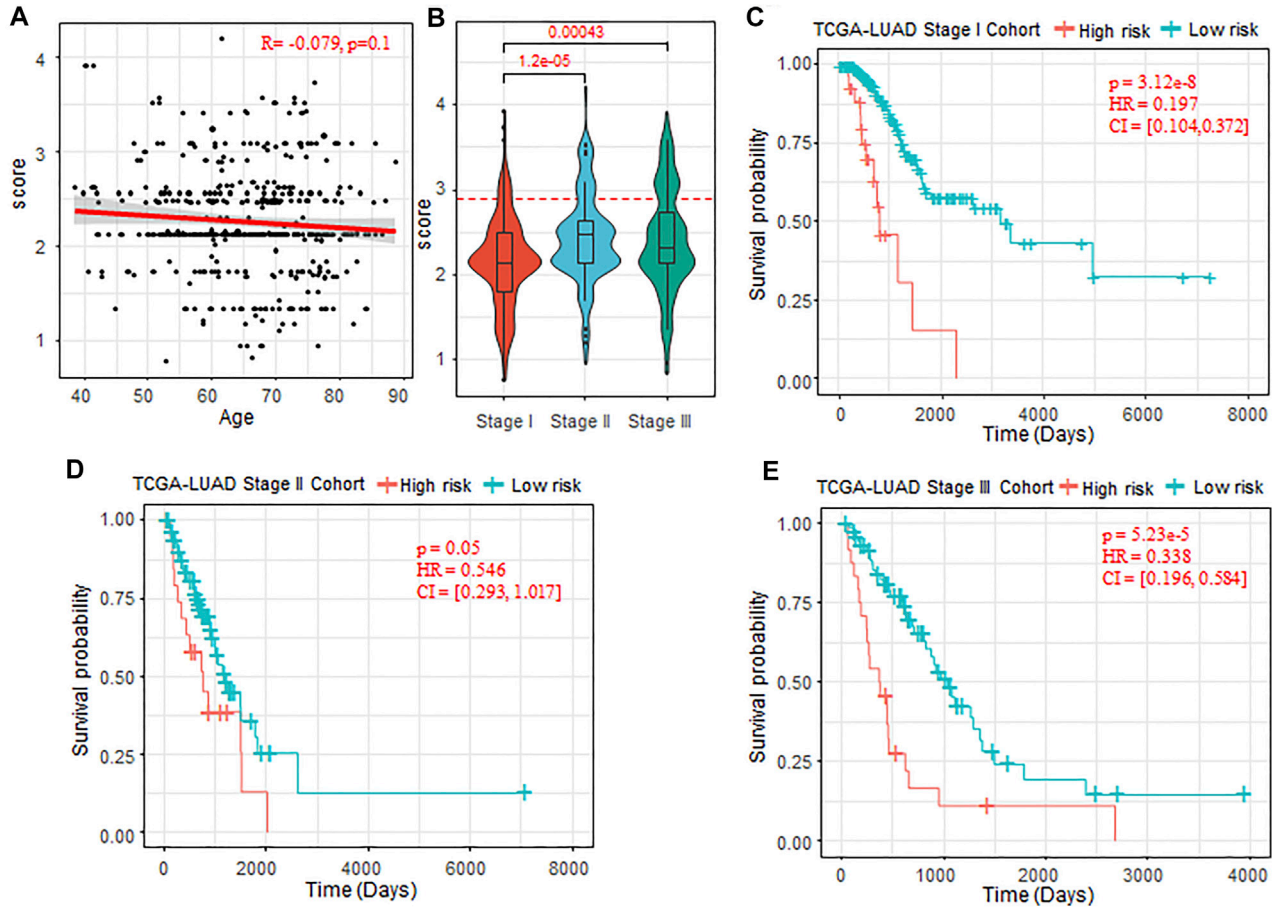


FIGURE 5 | Prognostic ability of the risk-score model with different clinical factors. **(A)** Correlation between the patient age and risk score predicted. **(B)** Comparison of risk score of patients in Stage I, Stage II, and Stage III. The Wilcoxon rank sum test was used. **(C–E)** Kaplan–Meier plots of OS in Stages I–III of TCGA-LUAD cohort when the risk-score cutoff set as 2.9. OS, overall survival; TCGA, The Cancer Genome Atlas; LUAD, lung adenocarcinoma.

METHODS

Data Preprocessing

The quantile normalization procedure is applied to the gene and miRNA expression separately and filter out the genes and miRNAs with the expression value 0 across more than 90% of the samples. We also applied the ComBat (Leek et al., 2012) to remove the batch effect between the data in TCGA dataset and CPTAC dataset. The DESeq2 (Love et al., 2014) was used to perform the differential expression analysis between the tumor and normal samples using the raw count data. Genes with Benjamini and Hochberg adjusted p-value of less than 1 e–3 and fold change larger than 2 were regarded as significantly DEGs.

Building the MicroRNA–Messenger RNA Negative Regulation Pairs

To obtain the relationship between miRNA and their target gene (mRNAs), we extracted the regulator factor miRNA of DEGs from three verified miRNA–target databases (miRecords (Xiao et al., 2009), miRTarBase (Huang et al., 2020b), and TarBase (Karagkouni et al., 2018)) using the "multiMiR" R package (Ru et al., 2014). These regulatory relationships were further refined based on the negative regulated relationship that one miRNA and its target genes were negatively related. Spearman's correlation test was applied to each miRNA–gene pair among 504 TP samples with both miRNA expression value and mRNA expression value available, and only the pairs with negative correlation coefficient and adjusted p-value < 0.01 remained.

MicroRNA Signature Selection

The procedure takes four steps to accomplish the miRNA signature selection. We first performed the functional enrichment analysis for the DEGs using the R/Biconductor package "clusterProfiler" (Yu et al., 2012), and functional terms with adjusted p-value of less than 0.05 were regarded as significantly enriched. We retained the miRNAs targeting the genes enriched in any functional terms. Next, we performed OS analysis for each of the remaining miRNAs, and the miRNAs with log-rank p-value of less than 0.05

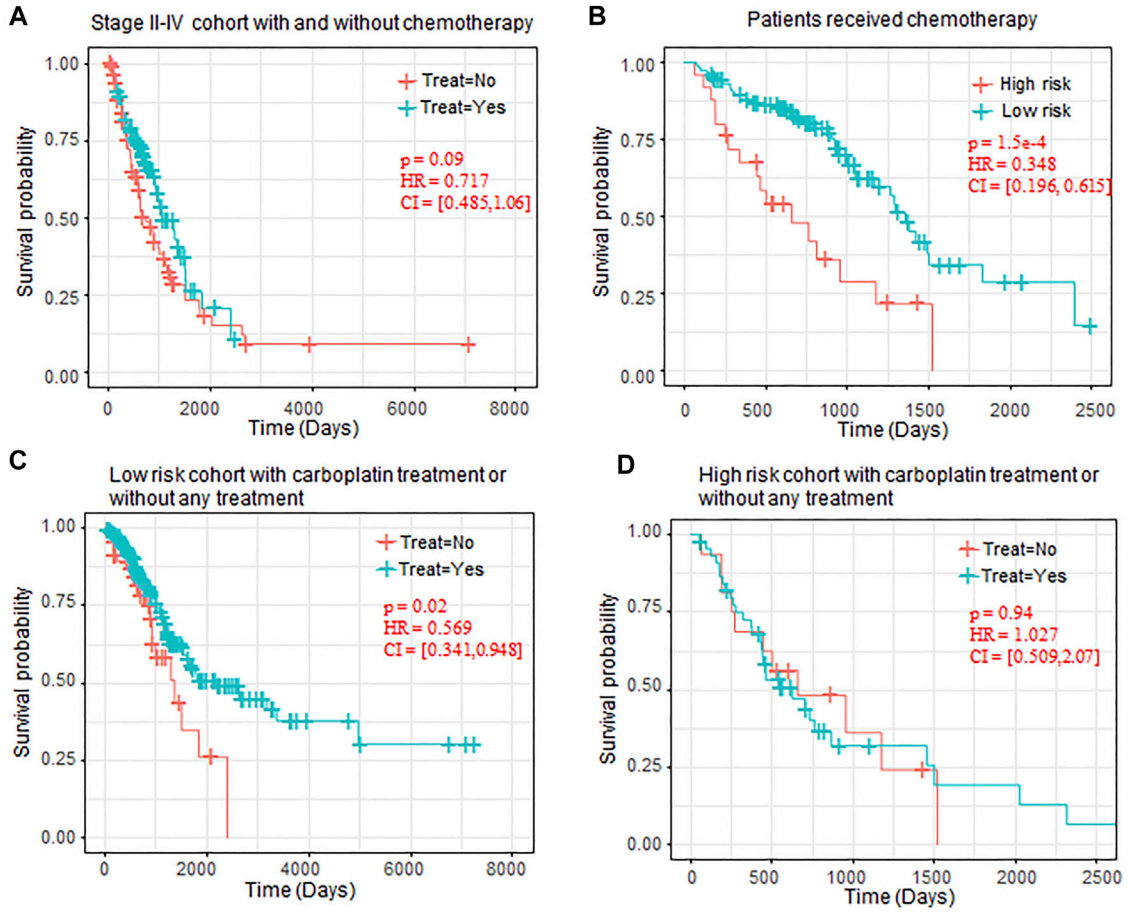


FIGURE 6 | OS comparison between patients with chemotherapy. **(A)** Kaplan–Meier plots of OS in Stage II–IV patients who received chemotherapy or not. **(B)** Kaplan–Meier plots of OS in high-risk and low-risk patients who received chemotherapy. **(C)** Kaplan–Meier plots of low-risk patients who received carboplatin or without any chemotherapy. **(D)** Kaplan–Meier plots of high-risk patients who received carboplatin or without any chemotherapy. OS, overall survival.

remained. To further refine the miRNA signatures, we evaluated the extent to which each miRNA contributes to predicting survival using the metric of variable importance using the vimp function from the R package "randomForestSRC" (Ishwaran et al., 2020). We calculated variable importance using random permutation of the variable approach. To ensure robustness, we repeated this step 5,000 times, and a rank matrix for the miRNAs was obtained based on the calculated variable importance. Using the rank matrix, we divided these miRNAs into three groups (including important miRNAs, secondary miRNA, and meaningless miRNAs) using R function hclust with the default parameters. The miRNAs regarded as important or secondary were selected as candidate miRNA signatures and ranked according to the median of the 5,000 ranks of the miRNA. Finally, we performed the multivariate survival analysis using the Cox regression model by feeding the candidate miRNA signatures in sequence. The miRNAs that reduced the prognostic ability of the model were excluded. Ultimately, the rest of the miRNAs were regarded as the signatures.

Building Risk-Score Estimator

For each miRNA signature, we calculated the optimal threshold that can divide the patients into the high-risk or low-risk group with the most significant OS time difference, and the beta (β) coefficient for each miRNA signature was also calculated with the optimal threshold. The risk score of a patient can be defined as follows:

$$Risk\ score\ =\ \sum_{i} s_i$$

and s_i represents the risk score for a certain miRNA i, which was calculated as follows:

$$s_i = \begin{cases} |\beta_i|,\ if\ \beta_i < 0 \text{ and miRNA expression lower than the related optimal threshold} \\ \beta,\ if\ \beta_i > 0 \text{ and miRNA expression higher than the related optimal threshold} \\ 0,\ else \end{cases}$$

Statistical Analysis

Time-dependent ROC curve and AUC were generated with R package "timeROC" (Blanche, 2015). Survival analysis and

univariate and multivariate Cox regression analyses were performed with R package "survival" (Therneau and Lumley, 2010). The Kaplan–Meier curves were plot with R package "survminer" (Kassambara et al., 2017). Heatmap was drawn with R package "pheatmap" (Kolde and Kolde, 2015). The *p*-values of each variable were corrected using the Benjamini and Hochberg (BH) method (Benjamini and Hochberg, 1995).

DISCUSSION

In this study, we have identified eight miRNA signatures associated with the OS of LUAD using both the miRNA expression and gene expression profiles obtained from TCGA-LUAD dataset. With these miRNA signatures, we built a novel risk-score model using both the optimal cutoff and corresponding beta coefficients; otherwise, the miRNA expression is used directly. This model divides LUAD patients into two groups (high-risk and low-risk) with significantly different OS times. The performance was proved to be consistent in both the training set (TCGA-LUAD) and independent validation set (CPTAC-LUAD).

Through consulting literature materials, we found that all the eight miRNAs were reported to be associated with various types of cancer, including lung cancer. Additionally, personalized cancer medicine is a clinical approach that strives to customize therapies based upon the genetic profiles of individual patient tumors. Our results further proved that stratification of LUAD patients is also important to the treatment and response to therapy. However, we also noted that the clinical information, such as treatment response, in TCGA database is mainly rough, and the results in this study need further investigation in the future.

Most importantly, as built based on the optimal threshold and corresponding beta coefficients, the proposed risk-score model was fit for different types of data, including both qualitative and quantitative. This risk-score model provided a new insight into the multi-omics data integration for prognosis.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JW, YL, and TS conceived the study. JW and Y-MM performed the algorithm development and downstream bioinformatics analysis. JW and YL wrote the manuscript. JX and TS revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China grants (Nos. 31801118, 31671377), Shanghai Municipal Science and Technology Major Project (Grant No. 2017SHZDZX01), Beihang University and Capital Medical University Plan (BHME-201904), the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, ECNU, and the Nurture projects for basic research of Shanghai Chest Hospital (No. 2020YNJCM06).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.741112/full#supplementary-material

REFERENCES

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bentwich, I. (2005). A Postulated Role for microRNA in Cellular Differentiation. *FASEB j.* 19, 875–879. doi:10.1096/fj.04-3609hyp

Blanche, P., Dartigues, J. F., and Jacqmin-Gadda, H. (2013). Estimating and Comparing Time-Dependent Areas Under Receiver Operating Characteristic Curves for Censored Event Times with Competing Risks. *Stat. Med.* 32, 5381–5397.

Cheng, A. M., Byrom, M. W., Shelton, J., and Ford, L. P. (2005). Antisense Inhibition of Human miRNAs and Indications for an Involvement of miRNA in Cell Growth and Apoptosis. *Nucleic Acids Res.* 33, 1290–1297. doi:10.1093/nar/gki200

Cheng, Y., Wang, K., Geng, L., Sun, J., Xu, W., Liu, D., et al. (2019). Identification of Candidate Diagnostic and Prognostic Biomarkers for Pancreatic Carcinoma. *Ebiomedicine* 40, 382–393. doi:10.1016/j.ebiom.2019.01.003

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). TCGAbiolinks: an R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res.* 44, e71. doi:10.1093/nar/gkv1507

Cui, X., Liu, Y., Sun, W., Ding, J., Bo, X., and Wang, H. (2020). Comprehensive Analysis of miRNA-Gene Regulatory Network with Clinical Significance in Human Cancers. *Sci. China Life Sci.* 63, 1201–1212. doi:10.1007/s11427-019-9667-0

Edwards, N. J., Oberti, M., Thangudu, R. R., Cai, S., McGarvey, P. B., Jacob, S., et al. (2015). The CPTAC Data Portal: A Resource for Cancer Proteomics Research. *J. Proteome Res.* 14, 2707–2713. doi:10.1021/pr501254j

Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010). Estimates of Worldwide burden of Cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* 127, 2893–2917. doi:10.1002/ijc.25516

Goldman, M., Craft, B., Zhu, J. C., and Haussler, D. (2017). The UCSC Xena System for Cancer Genomics Data Visualization and Interpretation. *Cancer Res.* 77, 2584. 10.1158/1538-7445.AM2017-2584.

Hess, J., Unger, K., Maihoefer, C., Schüttrumpf, L., Wintergerst, L., Heider, T., et al. (2019). A Five-MicroRNA Signature Predicts Survival and Disease Control of Patients with Head and Neck Cancer Negative for HPV Infection. *Clin. Cancer Res.* 25, 1505–1516. doi:10.1158/1078-0432.ccr-18-0776

Huang, H. Y., Lin, Y. C., Li, J., Huang, K. Y., Shrestha, S., Hong, H. C., et al. (2020). miRTarBase 2020: Updates to the Experimentally Validated microRNA-Target Interaction Database. *Nucleic Acids Res.* 48, D148–D154. doi:10.1093/nar/gkz896

Huang, S., Yang, J., Fong, S., and Zhao, Q. (2020). Artificial Intelligence in Cancer Diagnosis and Prognosis: Opportunities and Challenges. *Cancer Lett.* 471, 61–71. doi:10.1016/j.canlet.2019.12.007

Ishwaran, H., Kogalur, U. B., and Kogalur, M. U. B. (2020). *Package 'randomForestSRC*.

Karagkouni, D., Paraskevopoulou, M. D., Chatzopoulos, S., Vlachos, I. S., Tastsoglou, S., Kanellos, I., et al. (2018). DIANA-TarBase V8: a Decade-Long Collection of Experimentally Supported miRNA-Gene Interactions. *Nucleic Acids Res.* 46, D239–D245. doi:10.1093/nar/gkx1141

Kassambara, A., Kosinski, M., Biecek, P., Fabian, S., and survminer (2017). *Drawing Survival Curves Using'ggplot2*. R. Package Version 0.3 1.

Kolde, R., and Kolde, M. R. (2015). Package 'pheatmap'. *R. Package* 1, 790.

Lathwal, A., Kumar, R., Arora, C., and Raghava, G. P. S. (2020). Identification of Prognostic Biomarkers for Major Subtypes of Non-small-cell Lung Cancer Using Genomic and Clinical Data. *J. Cancer Res. Clin. Oncol.* 146, 2743–2752. doi:10.1007/s00432-020-03318-3

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* 28, 882–883. doi:10.1093/bioinformatics/bts034

Li, X., Shi, Y., Yin, Z., Xue, X., and Zhou, B. (2014). An Eight-miRNA Signature as a Potential Biomarker for Predicting Survival in Lung Adenocarcinoma. *J. Transl Med.* 12, 159. doi:10.1186/1479-5876-12-159

Li, X., Lu, C., Lu, Q., Li, C., Zhu, J., Zhao, T., et al. (2019). Differentiated Super-enhancers in Lung Cancer Cells. *Sci. China Life Sci.* 62, 1218–1228. doi:10.1007/s11427-018-9319-4

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Miska, E. A. (2007). Microrna Expression Profiles Classify Human Cancers. *Cytometry B-Clinical Cytometry* 72b, 126. doi:10.1002/cyto.b.v72b:6

Mlecnik, B., Bindea, G., Pagès, F., and Galon, J. (2011). Tumor Immunosurveillance in Human Cancers. *Cancer Metastasis Rev.* 30, 5–12. doi:10.1007/s10555-011-9270-7

Mounir, M., Lucchetta, M., Silva, T. C., Olsen, C., Bontempi, G., Chen, X., et al. (2019). New Functionalities in the TCGAbiolinks Package for the Study and Integration of Cancer Data from GDC and GTEx. *Plos Comput. Biol.* 15, e1006701. doi:10.1371/journal.pcbi.1006701

Novello, C., Pazzaglia, L., Cingolani, C., Conti, A., Quattrini, I., Manara, M. C., et al. (2013). miRNA Expression Profile in Human Osteosarcoma: Role of miR-1 and miR-133b in Proliferation and Cell Cycle Control. *Int. J. Oncol.* 42, 667–675. doi:10.3892/ijo.2012.1717

Ru, Y., Kechris, K. J., Tabakoff, B., Hoffman, P., Radcliffe, R. A., Bowler, R., et al. (2014). The multiMiR R Package and Database: Integration of microRNA-Target Interactions along with Their Disease and Drug Associations. *Nucleic Acids Res.* 42, e133. doi:10.1093/nar/gku631

Sheng, R., Li, X., Wang, Z., and Wang, X. (2020). Circular RNAs and Their Emerging Roles as Diagnostic and Prognostic Biomarkers in Ovarian Cancer. *Cancer Lett.* 473, 139–147. doi:10.1016/j.canlet.2019.12.043

Sherafatian, M., and Arjmand, F. (2019). Decision Tree-Based Classifiers for Lung Cancer Diagnosis and Subtyping Using TCGA miRNA Expression Data. *Oncol. Lett.* 18, 2125–2131. doi:10.3892/ol.2019.10462

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer Statistics, 2019. *CA A. Cancer J. Clin.* 69, 7–34. doi:10.3322/caac.21551

Sun, J., Wei, Q., Zhou, Y., Wang, J., Liu, Q., and Xu, H. (2017). A Systematic Analysis of FDA-Approved Anticancer Drugs. *BMC Syst. Biol.* 11, 87–43. doi:10.1186/s12918-017-0464-7

The Cancer Genome Atlas Research Network (2014). Comprehensive Molecular Profiling of Lung Adenocarcinoma. *Nature* 511, 543–550. doi:10.1038/nature13385

Therneau, T., and Lumley, T. (2010). Survival Analysis, Including Penalised Likelihood. *R. Package Version* 2, 36–14.

Volinia, S., Calin, G. A., Liu, C.-G., Ambs, S., Cimmino, A., Petrocca, F., et al. (2006). A microRNA Expression Signature of Human Solid Tumors Defines Cancer Gene Targets. *Pnas* 103, 2257–2261. doi:10.1073/pnas.0510565103

Wang, C., Ren, T., Wang, K., Zhang, S., Liu, S., Chen, H., et al. (2017). Identification of Long Non-coding RNA P34822 as a Potential Plasma Biomarker for the Diagnosis of Hepatocellular Carcinoma. *Sci. China Life Sci.* 60, 1047–1050. doi:10.1007/s11427-017-9054-y

Wang, J., Han, X., and Sun, Y. (2017). DNA Methylation Signatures in Circulating Cell-free DNA as Biomarkers for the Early Detection of Cancer. *Sci. China Life Sci.* 60, 356–362. doi:10.1007/s11427-016-0253-7

Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. (2009). miRecords: an Integrated Resource for microRNA-Target Interactions. *Nucleic Acids Res.* 37, D105–D110. doi:10.1093/nar/gkn851

Yang, Y., Yu, Y., and Lu, S. (2020). Effectiveness of PD-1/pd-L1 Inhibitors in the Treatment of Lung Cancer: Brightness and challenge. *Sci. China Life Sci.* 63, 1499–1514. doi:10.1007/s11427-019-1622-5

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Yu, S.-L., Chen, H.-Y., Chang, G.-C., Chen, C.-Y., Chen, H.-W., Singh, S., et al. (2008). MicroRNA Signature Predicts Survival and Relapse in Lung Cancer. *Cancer Cell* 13, 48–57. doi:10.1016/j.ccr.2007.12.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 23 November 2021
doi: 10.3389/fgene.2021.730141

The Construction of a Prognostic Model Based on a Peptidyl Prolyl *Cis–Trans* Isomerase Gene Signature in Hepatocellular Carcinoma

Huadi Shi†, *Fulan Zhong*†, *Xiaoqiong Yi, Zhenyi Shi, Feiyan Ou, Yufang Zuo* * *and Zumin Xu* *

Cancer Center, Affiliated Hospital of Guangdong Medical University, Zhanjiang, China

OPEN ACCESS

Edited by:
Yuriy L. Orlov, I.M.Sechenov First Moscow State Medical University, Russia

Reviewed by:
Dongjun Lee, Pusan National University, South Korea
Elvira Galieva, Novosibirsk State University, Russia
Liang Shi, Sun Yat-sen University, China

***Correspondence:**
Yufang Zuo
yufangzuo0102@163.com
Zumin Xu
zuminxu@gdmu.edu.cn

†*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *24 June 2021*
Accepted: *25 October 2021*
Published: *23 November 2021*

Citation:
Shi H, Zhong F, Yi X, Shi Z, Ou F, Zuo Y and Xu Z (2021) The Construction of a Prognostic Model Based on a Peptidyl Prolyl Cis–Trans *Isomerase Gene Signature in Hepatocellular Carcinoma. Front. Genet. 12:730141. doi: 10.3389/fgene.2021.730141*

Objective: The aim of the present study was to construct a prognostic model based on the peptidyl prolyl *cis–trans* isomerase gene signature and explore the prognostic value of this model in patients with hepatocellular carcinoma.

Methods: The transcriptome and clinical data of hepatocellular carcinoma patients were downloaded from The Cancer Genome Atlas and the International Cancer Genome Consortium database as the training set and validation set, respectively. Peptidyl prolyl *cis–trans* isomerase gene sets were obtained from the Molecular Signatures Database. The differential expression of peptidyl prolyl *cis–trans* isomerase genes was analyzed by R software. A prognostic model based on the peptidyl prolyl *cis–trans* isomerase signature was established by Cox, Lasso, and stepwise regression methods. Kaplan–Meier survival analysis was used to evaluate the prognostic value of the model and validate it with an independent external data. Finally, nomogram and calibration curves were developed in combination with clinical staging and risk score.

Results: Differential gene expression analysis of hepatocellular carcinoma and adjacent tissues showed that there were 16 upregulated genes. A prognostic model of hepatocellular carcinoma was constructed based on three gene signatures by Cox, Lasso, and stepwise regression analysis. The Kaplan–Meier curve showed that hepatocellular carcinoma patients in high-risk score group had a worse prognosis ($p < 0.05$). The receiver operating characteristic curve revealed that the area under curve values of predicting the survival rate at 1, 2, 3, 4, and 5 years were 0.725, 0.680, 0.644, 0.630, and 0.639, respectively. In addition, the evaluation results of the model by the validation set were basically consistent with those of the training set. A nomogram incorporating clinical stage and risk score was established, and the calibration curve matched well with the diagonal.

Conclusion: A prognostic model based on 3 peptidyl prolyl *cis–trans* isomerase gene signatures is expected to provide reference for prognostic risk stratification in patients with hepatocellular carcinoma.

Keywords: peptidyl prolyl *cis–trans* isomerase, nomogram, hepatocellular carcinoma, TCGA, ICGC

INTRODUCTION

The 2020 edition GLOBOCAN released by the World Health Organization shows that liver cancer ranks sixth in the number of new cases of malignant tumors worldwide and is the third leading cause of cancer death in the world (Sung et al., 2021). Hepatocellular carcinoma (HCC) is the most common pathological type of primary liver cancer, accounting for about 90% (Llovet et al., 2021). Current treatment options for HCC include radical hepatectomy, liver transplantation, arterial catheterization, radiotherapy, and chemotherapy. However, approximately 75% of patients are diagnosed with early disease after surgery relapse within 5 years. Moreover, surgical resection and liver transplantation are not appropriate for all HCC patients because most HCC patients are diagnosed as advanced or multifocal tumors, and the 5-year overall survival of HCC patients is less than 20% (Forner et al., 2018; Vibert et al., 2020; Yang and Heimbach, 2020). The TNM Classification of Malignant Tumors staging is one of the main reference indicators for prognosis assessment of HCC. However, TNM staging is insufficient in the assessment of prognosis due to the heterogeneity of tumors. The prognosis of HCC patients with the same TNM stage may vary, and even among HCC patients diagnosed with the same TNM stage and receiving similar clinical treatment, survival outcomes are various (Bruix et al., 2014; Dhir et al., 2016). Therefore, it is necessary to find more effective prognostic biomarkers in order to more accurately evaluate the prognosis and develop individualized treatment strategies.

Biological processes in the cell are extremely dynamic and complex events that are finely choreographed both spatially and temporally. The proper modulation of protein function is central to this orchestration. A number of regulatory mechanisms have been well-established, including post-translational chemical modifications of selected amino acid side chains, allosteric regulation, and regulated protein degradation (Lu et al., 2007). The peptidyl–prolyl *cis–trans* isomerases (PPIases) regulate the conversion between cis and trans conformations of proteins as a molecular timer and play an important regulatory role in the process of life activities (Lu et al., 2007). The PPIase superfamily comprises four structurally unrelated families: cyclophilins, FK506-binding proteins, parvulins, and the protein phosphatase 2A phosphatase activator. These proteins exhibit well-conserved CYP or FKBP domains. These four subfamilies of PPIases are not similar in their sequences and three-dimensional structures, but these proteins exhibit well-conserved CYP or FKBP domains and can all catalyze the *cis–trans* isomerism of the peptide–proline amide bond (Fischer et al., 1989; Fanghänel and Fischer, 2004; Jordens et al., 2006; Mueller and Bayer, 2008). Many members of the PPIases gene family have recently been found to be closely associated with cancer progression and prognosis (Hojo et al., 1999; Bao et al., 2004; Ni et al., 2010; Annett et al., 2020). However, the prognostic value of the PPIase gene signature in HCC remains unclear.

In this study, the transcriptome and clinical data of HCC patients were downloaded from TCGA and ICGC databases as training set and validation set, respectively. A prognostic model based on the PPIase gene signature was established by using Cox, Lasso, and stepwise regression methods. Kaplan–Meier survival analysis was used to evaluate the prognostic value of the model and validate it with an independent external data. In addition, nomogram and calibration curves were developed in combination with clinical staging and risk score.

METHODS

Acquisition of Peptidyl Prolyl *Cis–Trans* Isomerase Gene Sets

We obtained 43 PPIase genes from the GO_PROTEIN_PEPTIDYL_PROLYL_ISOMERIZATION gene sets in the Molecular Signatures Database (MSigDB v7.2, http//: software.broadinstitute.org/gsea/msigdb) (Liberzon et al., 2011).

Transcriptome and Clinical Data of Hepatocellular Carcinoma

Transcriptome and clinical data were downloaded from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) (Blum et al., 2018). Information on the gene expression and comparing clinical data (377 cases; data format: BCR XML) were downloaded from the level 3 gene expression information (FPKM normalized) of the TCGA LIHC cohort. The data from TCGA were used as the training set, and the data from ICGC were used as the validation set. Another RNA-seq dataset of 240 primary HCC patients together with corresponding clinical information was accessed from the ICGC (https://dcc.icgc.org/, LIRI-JP) (Kennedy et al., 2012), which was used as a cohort for external validation of the signature. The clinicopathological data collected included sex, age, stage, grade, survival status, and survival duration in days. Our study was in accordance with the publication guidelines provided by TCGA.

Identification of Differentially Expressed Genes

The differential expression of the PPIase gene in 370 HCC tissues and 50 para-cancerous tissues was analyzed by the "limma" package of R 3.6.1 software. The criteria for selection of differentially expressed genes were FDR <0.05, $|\log2FC| \geq 1$, FDR: false discovery rate, and FC: fold change.

Construction of the Prognostic Risk Score Model

The clinical data of HCC were merged with the expression data of PPIase genes. The "survival" package of R software was used to perform univariate Cox regression analysis. The hazard ratio (HR) and corresponding p value of each PPIase gene were obtained by univariate Cox regression analysis. When the p value was less than 0.05, the gene was selected for further analysis. In order to reduce the collinearity between genes and prevent the over-fitting of prognostic risk model variables, Lasso regression was used to further analyze the variables obtained from univariate Cox regression (Tibshirani, 1997). Subsequently, we

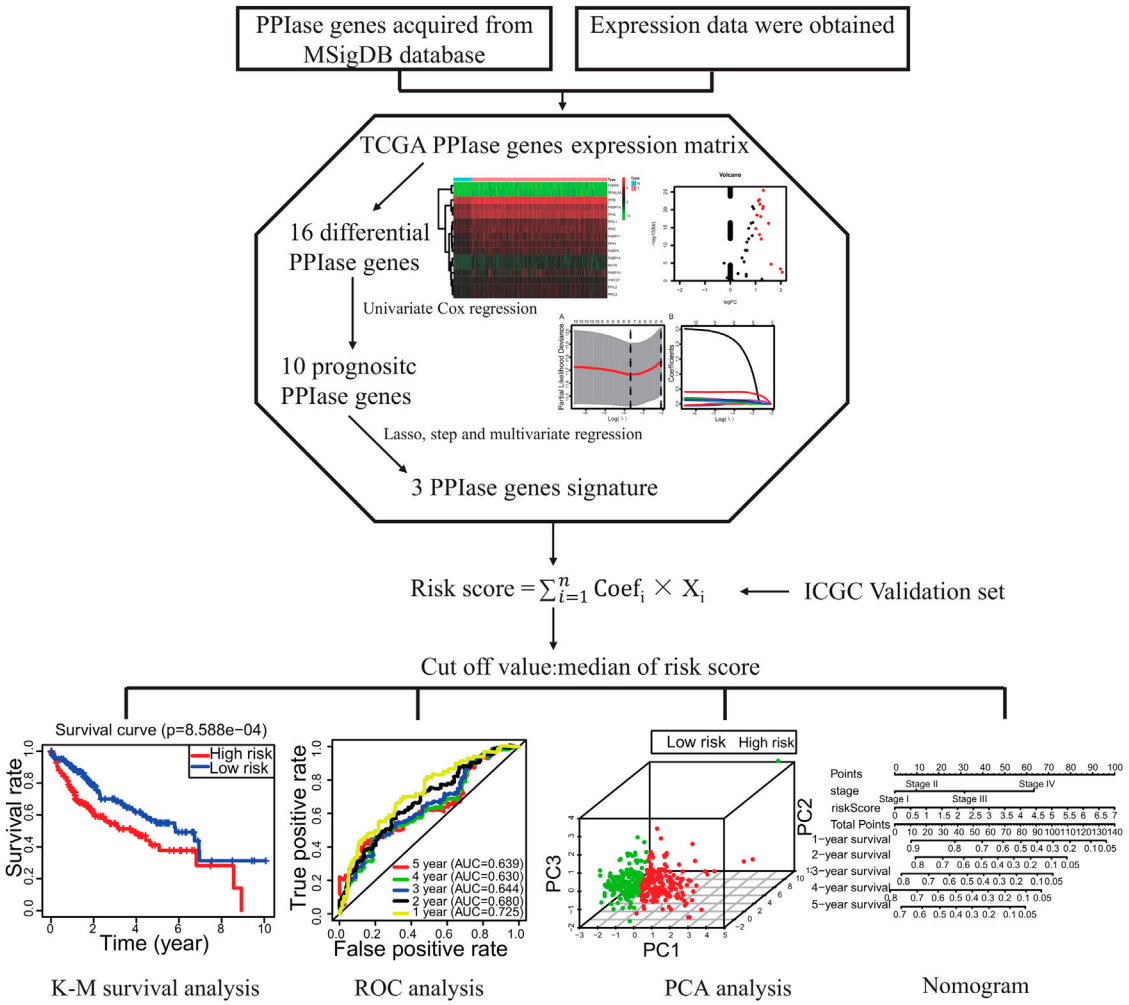


FIGURE 1 | Flowchart for profiling the peptidyl prolyl *cis*–trans isomerase gene signatures.

performed further variable filtering through the "step" function, which was a stepwise regression analysis based on AIC information statistics. In addition, the coefficient of each PPIase gene was calculated by multivariate Cox regression analysis. Finally, the risk score equation was constructed as follows:

$$risk\ core = \sum_{i=1}^{n} \mathrm{Coef}_i \times \mathrm{X}_i,$$

where Coef is the coefficient, n is the number of genes, X is the expression value of the gene, and i refers to the serial number.

Evaluation and Validation of the Prognostic Risk Score Model

The risk score of each HCC patient was calculated by the risk score equation. Patients were divided into low-risk and high-risk groups according to the median of risk score as the cutoff value. Kaplan–Meier survival analysis was performed using the "survival" package of R software. The "timeROC" package was used to draw the ROC curve of the model. The area under curve (AUC) was calculated to evaluate the sensitivity and specificity of the prognostic model. Principal component analysis was performed to explore the distribution pattern of high- and low-risk groups according to PPIase gene expression. In addition, we performed univariate and multivariate Cox regression analyses to investigate whether the risk score can be an independent predictor of overall survival in HCC patients. Covariates included age, stage, and grade. To verify the reliability of the model, we downloaded the LIRI-JP dataset from the ICGC database as the validation set. The risk score for each patient was calculated using the same formula as the training set.

The Construction of Nomogram and Calibration Curves

In order to better evaluate the clinical significance of the model and facilitate clinical application, a nomogram integrating TNM staging and prognostic risk score was constructed. Clinicians can

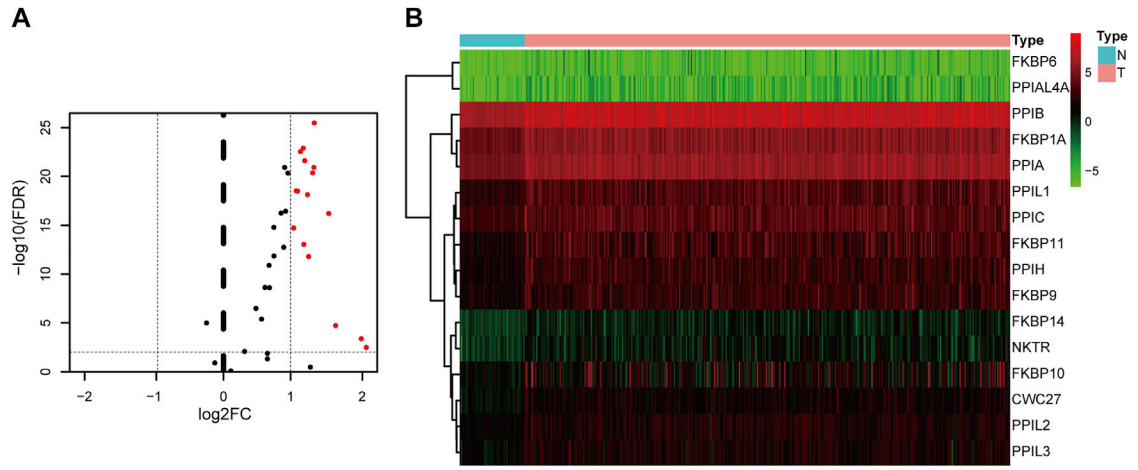


FIGURE 2 | Differentially expressed PPIase genes. Volcano plot **(A)** and heat map **(B)** of differential expression in hepatocellular carcinoma and para-cancerous tissues.

TABLE 1 | Prognostic values of 16 PPIase genes.

Gene name	Regression coefficient	Hazard ratio (95%confidence interval)	*p*-value
FKBP6	3.227	25.2 (2.7–234.7)	**0.005**
CWC27	0.892	2.4 (1.6–3.6)	**0.000**
PPIB	0.010	1.0 (0.7–1.4)	0.950
FKBP11	0.152	1.2 (1.0–1.4)	0.118
PPIH	0.517	1.7 (1.3–2.2)	**0.000**
PPIAL4A	-0.446	0.6 (0.0–21)	0.802
FKBP10	0.125	1.1 (1.0–1.3)	**0.045**
PPIL2	0.372	1.5 (1.0–2.1)	**0.049**
FKBP1A	0.503	1.7 (1.3–2.1)	**0.000**
PPIL1	0.492	1.6 (1.2–2.1)	**0.000**
FKBP9	0.289	1.3 (1.0–1.7)	**0.022**
FKBP14	0.379	1.5 (1.0–2.1)	**0.031**
PPIA	0.582	1.8 (1.3–2.4)	**0.000**
PPIC	0.108	1.1 (0.9–1.4)	0.398
NKTR	0.094	1.1 (0.8–1.5)	0.539
PPIL3	0.002	1.0 (0.7–1.4)	0.993

Abbreviation: PPIase, peptidyl prolyl cis–*trans isomerase. Bold values indicate* p < *0.05.*

quantitatively assess the prognostic risk based on the score for each risk variable in the model. Finally, the calibration curve was drawn to evaluate the accuracy of the nomogram.

RESULTS

Differential Expression of Peptidyl Prolyl *Cis–Trans* Isomerase Genes

Flowchart for profiling the PPIase genes of HCC (**Figure 1**). The TCGA–LIHC data were downloaded. There were 50 para-cancerous tissues and 370 HCC tissues which were included after data collection. The results showed that compared with the para-cancerous tissues, there were 16 upregulated PPIase genes in HCC tissues (**Figures 2A,B**).

Construction of the Prognostic Risk Score Model Based on 3 Genes

Sixteen differentially expressed PPIase genes were included in univariate Cox regression analysis. There were 10 genes associated with survival in HCC patients, including FKBP6, CWC27, PPIH, FKBP10, PPIL2, FKBP1A, PPIL1, FKBP9, FKBP14, and PPIA (**Table 1**, $p < 0.05$). Lasso regression was applied to further screen the 10 prognostic PPIase genes, in order to reduce the influence of collinearity among genes and prevent over-fitting of risk model variables constructed later. The results of Lasso regression were included in the 7 PPIase genes: FKBP6, CWC27, PPIH, FKBP1A, PPIL1, FKBP14, and PPIA (**Figure 3**). Finally, a prognostic model based on the mRNA expression and coefficients of the 3 genes was finally obtained by multivariate Cox and stepwise regression analyses. The coefficients of each

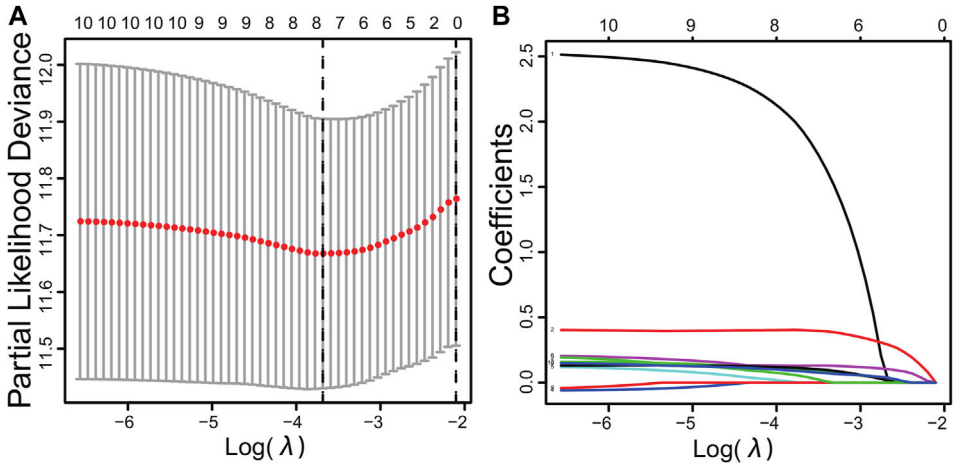


FIGURE 3 | Cross-validation results **(A)** and dynamic process diagram of Lasso regression screening variables **(B)**.

TABLE 2 | Most prognosis-related PPIase genes.

Gene name	Coefficient	Hazard ratio (95%confidence interval)	*p*-value
FKBP6	2.679,926	14.58 (1.36–156)	0.027
CWC27	0.669,348	1.953 (1.28–2.99)	0.002
FKBP1A	0.306,267	1.358 (1.01–1.83)	0.045

Abbreviation: PPIases, peptidyl prolyl cis–trans *isomerases.*

gene are listed in **Table 2**. The risk score was quantified by the following formula:

$$\text{risk score} = (2.68 \times \text{FKBP6}) + (0.67 \times \text{CWC27}) + (0.31 \times \text{FKBP1A}).$$

Evaluation of the Peptidyl Prolyl *Cis–Trans* Isomerase Gene Signature Model

The Kaplan–Meier curve showed that HCC patients in high-risk score group had a worse prognosis ($p < 0.05$, **Figure 4A**). The ROC curve revealed that the AUC values of predicting survival rate at 1, 2, 3, 4, and 5 years were 0.725, 0.680, 0.644, 0.630, and 0.639, respectively (**Figure 4B**). The results of principal component analysis revealed that there were significant differences in the distribution patterns of HCC in the high-risk and low-risk groups (**Figure 4C**).

Validation of the Peptidyl Prolyl *Cis-Trans* Isomerase Gene Signature Model

In order to verify the reliability of the model, we applied the external dataset from the ICGC database for validation. There were 230 HCC tissues which were included after data collation. The patients of the validation set were divided into high-risk ($n = 115$) and low-risk groups ($n = 115$) based on the median of risk score. Consistent with the results of the TCGA dataset, the Kaplan–Meier curve showed that HCC patients in high-risk score group had a worse prognosis ($p < 0.05$, **Figure 4D**). The ROC curve revealed that the AUC values of predicting survival rate at 1, 2, 3, 4, and 5 years were 0.601, 0.679, 0.67, 0.688, and 0.688, respectively (**Figure 4E**). The results of principal component analysis revealed that there were significant differences in the distribution patterns of HCC in the high-risk and low-risk groups (**Figure 4F**). It is suggested that the model has a good inclusiveness.

Risk Score as an Independent Prognostic Factor

Univariate and multivariate Cox regression analyses were performed to investigate whether the risk score could be an independent predictor of prognosis in patients with HCC. Univariate Cox regression analysis showed a significant correlation between the risk score and overall survival in the training set (HR = 1.602, 95% CI = 1.346–1.908, $p < 0.001$, **Figure 5A**). Multivariate Cox analysis suggested that the risk score was an independent prognostic predictor (HR = 1.475, 95% CI = 1.194–1.821, $p < 0.001$, **Figure 5B**). Similarly, univariate Cox regression analysis revealed that the risk score was related to overall survival in the validation set (HR = 1.375, 95% CI = 1.164–1.583, $p < 0.001$, **Figure 5C**). Multivariate Cox analysis suggested that the risk score was an independent prognostic predictor in the validation set (HR = 1.277, 95% CI = 1.070–1.524, $p = 0.007$, **Figure 5D**).

The Construction of Nomogram and Calibration Curves

The nomogram is a clinical tool that allows clinicians to determine the prognosis of patients by adding the score of each risk variable in the model to obtain the total score and the corresponding survival rate. Therefore, this study constructed a nomogram combining TNM staging and risk score. The ROC curve showed that the AUC value of risk score predicted 1-year

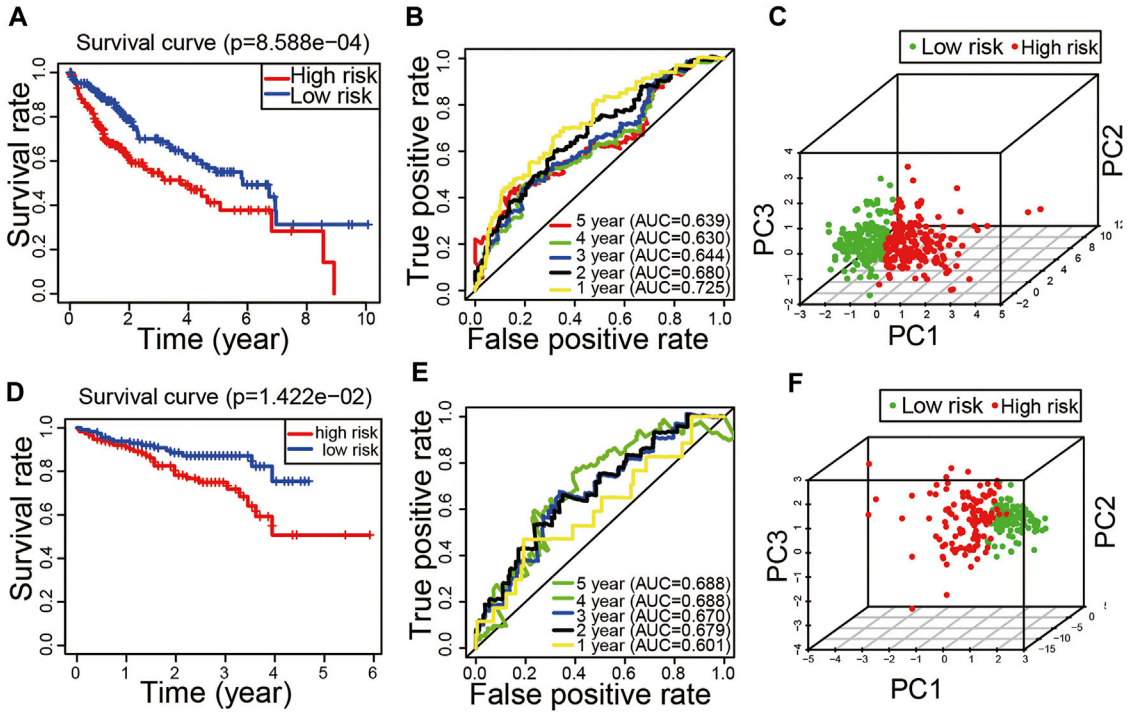


FIGURE 4 | Evaluation and validation of the peptidyl prolyl cis–trans isomerase gene signature model. **(A)** Kaplan–Meier curve in the training set. **(B)** Time-dependent receiver operating characteristic curve in the training set. **(C)** Principal component analysis in the training set. **(D)** Kaplan–Meier curve in the validation set. **(E)** Time-dependent receiver operating characteristic curve in the validation set. **(F)** Principal component analysis in the validation set.

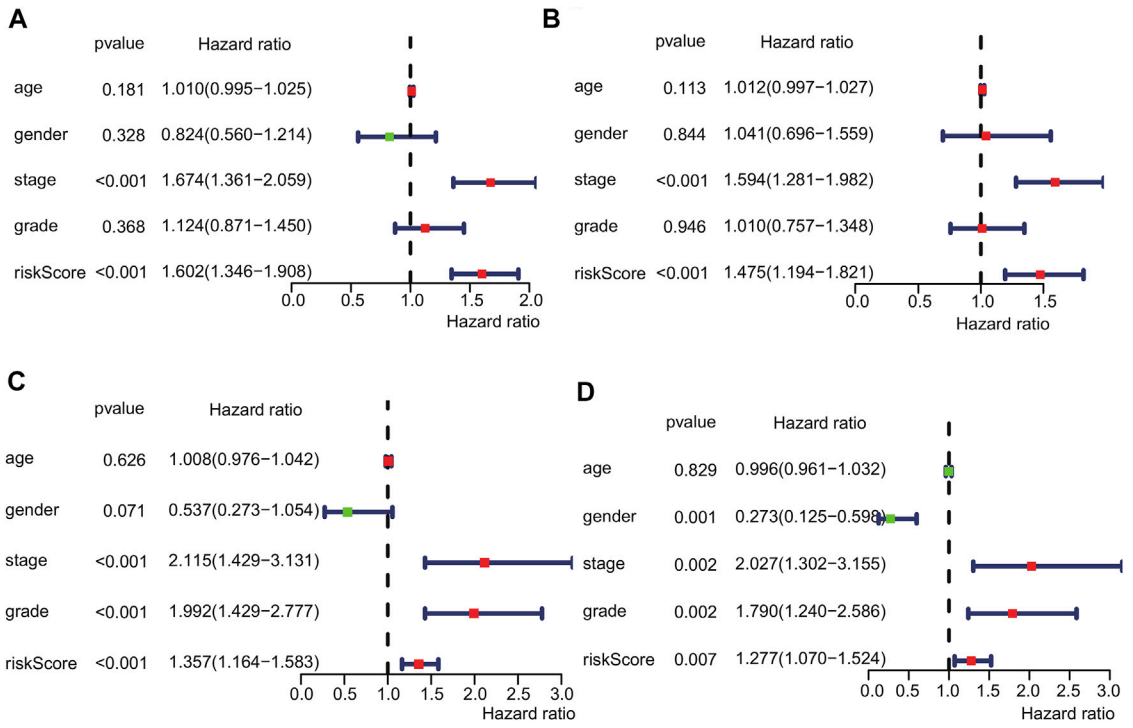


FIGURE 5 | Risk score as an independent prognostic factor. Univariate **(A)** and multivariate **(B)** Cox regression analyses in the training set. Univariate **(C)** and multivariate **(D)** Cox regression analyses in the validation set.

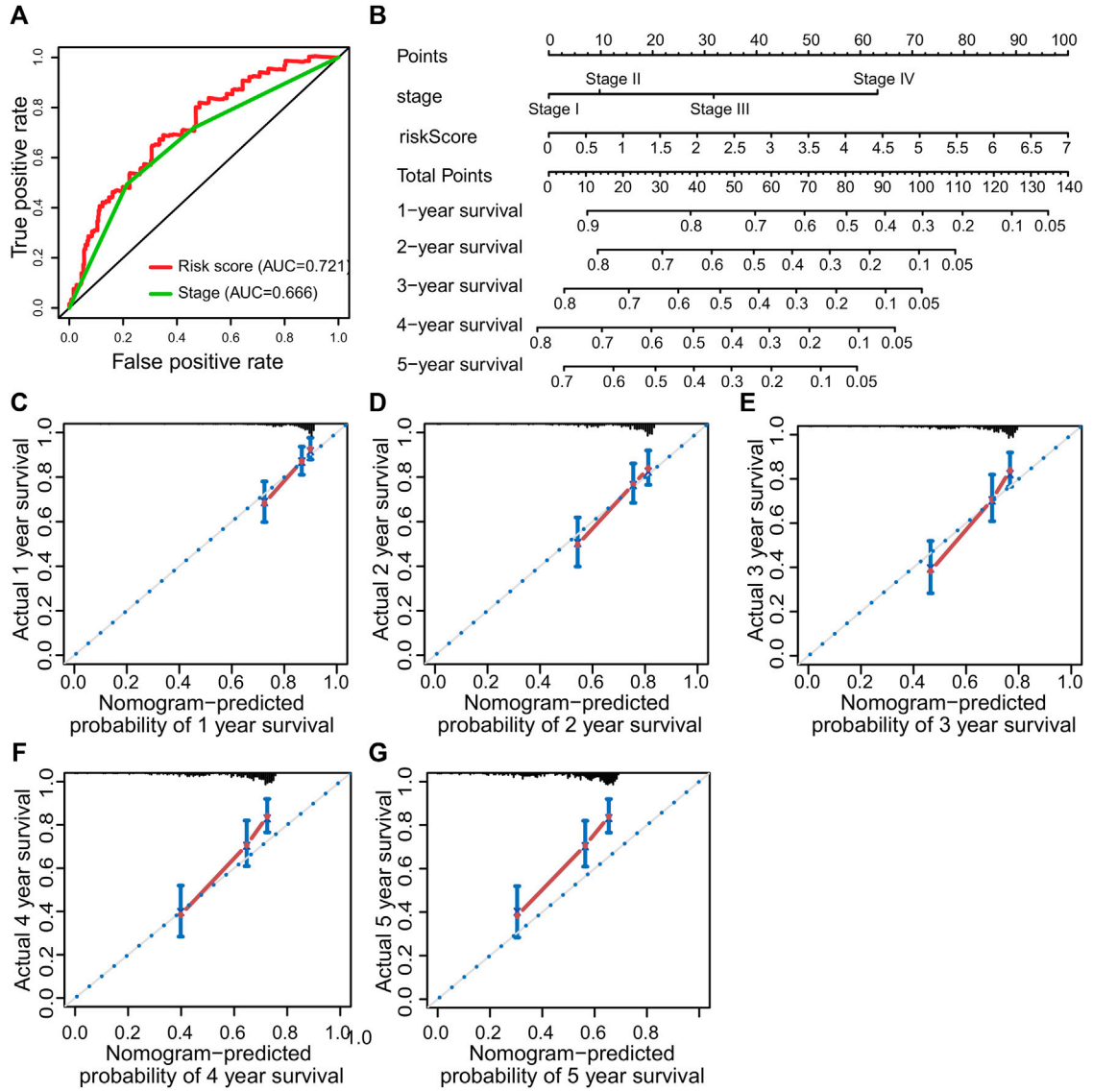


FIGURE 6 | Prognostic nomogram was constructed to predict the overall survival probability based on the training set of patients with hepatocellular carcinoma. **(A)** ROC curve analysis. **(B)** The nomogram to predict 1-, 2-, 3-, 4- and 5-year OS overall survival in the train set. **(C–G)** The calibration plots for predicting patient 1-, 2-, 3-, 4- and 5-year OS, respectively.

survival was greater than stage (**Figure 6A**). The nomogram revealed that the risk score was the most important factor among the various clinical parameters (**Figure 6B**). Moreover, the 1-year, 2-year, 3-year, 4-year, and 5-year calibration curves have a high matching degree with the diagonal (**Figures 6C–G**).

DISCUSSION

In recent years, with the rapid development of next-generation sequencing technology and precision medicine, more and more evidence indicates that gene signatures of mRNA level have good potential in predicting the prognosis of many cancers, including HCC. For example, the application of bioinformatics methods to construct a prognostic model based on the gene signature of autophagy, M6A methylation, and immunity have been reported for a variety of cancers, which is even better than TNM staging to a certain extent (Brebi et al., 2014; Frost and Amos, 2017; Liu et al., 2021). However, most of the existing signatures have not been widely used in clinical practice of HCC because the reliability of models is affected by many factors such as over-fitting. In order to prevent over-fitting, some recent studies have adopted the regularization method, and the model has good reliability (Wang et al., 2020; Wang et al., 2021a; Li et al., 2021). Therefore, this study intends to use a combination of multiple regularization methods to construct an HCC prognosis model based on the PPIase gene set.

Many members of the PPIases gene family have recently been found to be closely associated with cancer progression

and prognosis (Hojo et al., 1999; Bao et al., 2004; Ni et al., 2010; Annett et al., 2020). Therefore, we attempted to construct a prognostic model using the PPIases gene set. Surprisingly, we found that a model based on 3 PPIases gene signatures had good prognostic value. Multivariate Cox analysis suggested that the risk score was an independent prognostic predictor. The Kaplan–Meier curve showed that the prognosis of HCC patients in the high-risk group was worse. The AUC value of the ROC curve for predicting 1-year survival was greater than 0.7. A useful line nomogram was also successfully constructed.

In this study, PPIase gene differential expression was analyzed in HCC and adjacent tissues. Finally, 3 genes (FKBP6, CWC27, and FKBP1A) most related to prognosis were screened out by Cox and Lasso regression methods. It was reported that promoter methylation of FKBP6 can be used as a biomarker for the diagnosis of cervical cancer (Fischer et al., 1989). Another research showed that CWC27 can be used as a biomarker for the prognosis of bladder cancer (Wan et al., 2020). FKBP1A has also been reported to play a role in promoting tumor progression. Zhang et al. (2019) found that FKBP1A affected the proliferation and migration of prostate cancer cells (Lipunova et al., 2019). Romano et al. (2008) found that knockdown of FKBP1A can activate the TGF-β signaling pathway in chronic lymphocytic leukemia cells (Zhang et al., 2019). These studies suggest that FKBP1A may play a role in promoting cancer development in chronic lymphocytic leukemia and prostate cancer.

The nomogram is a clinical tool that allows clinicians to determine the prognosis of patients by adding the score of each risk variable in the model to obtain the total score and the corresponding survival rate (Romano et al., 2008). In recent years, the nomogram has been widely used as one of the practical tools in the assessment of cancer prognosis (Ohori Tatsuo Gondo And Riu Hamada et al., 2009; Zhou et al., 2021a; Wang et al., 2021b; Wu et al., 2021). Calibration curves are often used to evaluate the accuracy of a nomogram. The calibration curves of an ideal model just fall on the diagonal, and the more the calibration curves match the diagonal, the higher will be the prediction accuracy (Iasonos et al., 2008; Zhou et al., 2021b). As shown in **Figures 5B,C,D**, the calibration curve for predicting the survival rate at 1, 2, and 3 years has a good matching degree with the diagonal, suggesting a high accuracy of the model. Our model might provide a new reference for prognostic risk stratification assessment in HCC patients.

However, our model also has some limitations. First, further studies with additional external datasets are needed to confirm the prognostic value of the model. Second, the prognostic model is based on retrospective data, which is prone to selection bias, information bias, and other biases.

CONCLUSION

A prognostic model based on 3 PPIase gene signatures is expected to provide reference for prognostic risk stratification in patients with HCC.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

AUTHOR CONTRIBUTIONS

HS and FZ were responsible for original draft writing, acquisition, investigation, conceptualization, visualization, and software. XY, ZS, and FO were involved in methodology. ZX and YZ were involved in the conception and design of the study and revised the manuscript. All authors have read and agreed to the final version of the manuscript.

FUNDING

This work was supported by the Natural Science Foundation of Guangdong Province (No. 2020A1515010048), the Traditional Chinese Medicine research projects of Traditional Chinese Medicine Bureau of Guangdong Province (No. 20201179), and the non-funded science and technology projects of Zhanjiang city (No. 2019B01021).

REFERENCES

Annett, S., Moore, G., and Robson, T. (2020). FK506 Binding Proteins and Inflammation Related Signalling Pathways; Basic Biology, Current Status and Future Prospects for Pharmacological Intervention. *Pharmacol. Ther.* 215, 107623. doi:10.1016/j.pharmthera.2020.107623

Bao, L., Kimzey, A., Sauter, G., Sowadski, J. M., Lu, K. P., and Wang, D.-G. (2004). Prevalent Overexpression of Prolyl Isomerase Pin1 in Human Cancers. *Am. J. Pathol.* 164 (5), 1727–1737. doi:10.1016/S0002-9440(10)63731-5

Blum, A., Wang, P., and Zenklusen, J. C. (2018). SnapShot: TCGA-Analyzed Tumors. *Cell* 173 (2), 530. doi:10.1016/j.cell.2018.03.059

Brebi, P., Maldonado, L., Noordhuis, M. G., Ili, C., Leal, P., Garcia, P., et al. (2014). Genome-wide Methylation Profiling Reveals Zinc finger Protein 516 (ZNF516) and FK-506-Binding Protein 6 (FKBP6) Promoters Frequently Methylated in Cervical Neoplasia, Associated with HPV Status and Ethnicity in a Chilean Population. *Epigenetics* 9 (2), 308–317. doi:10.4161/epi.27120

Bruix, J., Gores, G. J., and Mazzaferro, V. (2014). Hepatocellular Carcinoma: Clinical Frontiers and Perspectives. *Gut* 63 (5), 844–855. doi:10.1136/gutjnl-2013-306627

Dhir, M., Melin, A. A., Douaiher, J., Lin, C., Zhen, W., Hussain, S. M., et al. (2016). A Review and Update of Treatment Options and Controversies in the Management of Hepatocellular Carcinoma. *Ann. Surg.* 263 (6), 1112–1125. doi:10.1097/SLA.0000000000001556

Fanghänel, J. r., and Fischer, G. (2004). Insights into the Catalytic Mechanism of Peptidyl Prolyl Cis/trans Isomerases. *Front. Biosci.* 9, 3453–3478. doi:10.2741/1494

Fischer, G., Wittmann-Liebold, B., Lang, K., Kiefhaber, T., and Schmid, F. X. (1989). Cyclophilin and Peptidyl-Prolyl Cis-Trans Isomerase Are Probably Identical Proteins. *Nature* 337 (6206), 476–478. doi:10.1038/337476a0

Forner, A., Reig, M., and Bruix, J. (2018). Hepatocellular Carcinoma. *The Lancet* 391 (10127), 1301–1314. doi:10.1016/S0140-6736(18)30010-2

Frost, H. R., and Amos, C. I. (2017). Gene Set Selection via LASSO Penalized Regression (SLPR). *Nucleic Acids Res.* 45 (12), e114. doi:10.1093/nar/gkx291

Hojo, M., Morimoto, T., Maluccio, M., Asano, T., Morimoto, K., Lagman, M., et al. (1999). Cyclosporine Induces Cancer Progression by a Cell-Autonomous Mechanism. *Nature* 397 (6719), 530–534. doi:10.1038/17401

Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How to Build and Interpret a Nomogram for Cancer Prognosis. *Jco* 26 (8), 1364–1370. doi:10.1200/JCO.2007.12.9791

Jordens, J., Janssens, V., Longin, S., Stevens, I., Martens, E., Bultynck, G., et al. (2006). The Protein Phosphatase 2A Phosphatase Activator Is a Novel Peptidyl-Prolyl Cis/trans-Isomerase. *J. Biol. Chem.* 281 (10), 6349–6357. doi:10.1074/jbc.M507760200

Kennedy, K., Troncoso, A., Kaan, T., Rial-Sebbag, E., and Aa, V. V. (2012). Legal Aspects of Genetic Databases for International Biomedical Research: the Example of the International Cancer Genome Consortium (ICGC). *Rev. Derecho Genoma Hum.* 2012 (37), 15–34.

Li, W., Liu, J., Ma, Z., Zhai, X., Cheng, B., and Zhao, H. (2021). m6A RNA Methylation Regulators Elicit Malignant Progression and Predict Clinical Outcome in Hepatocellular Carcinoma. *Dis. Markers* 2021, 18859590–18859602. doi:10.1155/2021/8859590

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics* 27 (12), 1739–1740. doi:10.1093/bioinformatics/btr260

Lipunova, N., Wesselius, A., Cheng, K. K., van Schooten, F. J., Cazier, J.-B., Bryan, R. T., et al. (2019). External Replication of Urinary Bladder Cancer Prognostic Polymorphisms in the uk Biobank. *Front. Oncol.* 9, 1082. doi:10.3389/fonc.2019.01082

Liu, F., Yang, Z., Zheng, L., Shao, W., Cui, X., Wang, Y., et al. (2021). A Tumor Progression Related 7-Gene Signature Indicates Prognosis and Tumor Immune Characteristics of Gastric Cancer. *Front. Oncol.* 11, 690129. doi:10.3389/fonc.2021.690129

Llovet, J. M., Kelley, R. K., Villanueva, A., Singal, A. G., Pikarsky, E., Roayaie, S., et al. (2021). Hepatocellular Carcinoma. *Nat. Rev. Dis. Primers* 7 (1), 6. doi:10.1038/s41572-020-00240-3

Lu, K. P., Finn, G., Lee, T. H., and Nicholson, L. K. (2007). Prolyl Cis-Trans Isomerization as a Molecular Timer. *Nat. Chem. Biol.* 3 (10), 619–629. doi:10.1038/nchembio.2007.35

Mueller, J. W., and Bayer, P. (2008). Small Family with Key Contacts: Par14 and Par17 Parvulin Proteins, Relatives of Pin1, Now Emerge in Biomedical Research. *Perspect. Medicin Chem.* 2, PMC.S496–20. doi:10.4137/pmc.s496

Ni, L., Yang, C.-S., Gioeli, D., Frierson, H., Toft, D. O., and Paschal, B. M. (2010). FKBP51 Promotes Assembly of the Hsp90 Chaperone Complex and Regulates Androgen Receptor Signaling in Prostate Cancer Cells. *Mol. Cel Biol* 30 (5), 1243–1253. doi:10.1128/MCB.01891-08

Ohori Tatsuo Gondo And Riu Hamada, M., Gondo, T., Hamada, R., and Hamada, R. (2009). Nomogram as Predictive Model in Clinical Practice. *Gan To Kagaku Ryoho* 36 (6), 901–906.

Romano, S., Mallardo, M., Chiurazzi, F., Bisogni, R., D'Angelillo, A., Liuzzi, R., et al. (2008). The Effect of FK506 on Transforming Growth Factor Signaling and Apoptosis in Chronic Lymphocytic Leukemia B Cells. *Haematologica* 93 (7), 1039–1048. doi:10.3324/haematol.12402

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: Globocan Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Tibshirani, R. (1997). The Lasso Method for Variable Selection in the Cox Model. *Statist. Med.* 16 (4), 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

Vibert, E., Schwartz, M., and Olthoff, K. M. (2020). Advances in Resection and Transplantation for Hepatocellular Carcinoma. *J. Hepatol.* 72 (2), 262–276. doi:10.1016/j.jhep.2019.11.017

Wan, Q., Jin, L., Su, Y., Liu, Y., Li, C., and Wang, Z. (2020). Development and Validation of Autophagy-related-gene Biomarker and Nomogram for Predicting the Survival of Cutaneous Melanoma. *IUBMB Life* 72 (7), 1364–1378. doi:10.1002/iub.2258

Wang, C., Qiu, J., Chen, S., Li, Y., Hu, H., Cai, Y., et al. (2021). Prognostic Model and Nomogram Construction Based on Autophagy Signatures in Lower Grade Glioma. *J. Cel Physiol* 236 (1), 235–248. doi:10.1002/jcp.29837

Wang, H., Ma, X., Liu, J., Wan, Y., Jiang, Y., Xia, Y., et al. (2020). Prognostic Value of an Autophagy-Related Gene Expression Signature for Endometrial Cancer Patients. *Cancer Cel Int* 20, 306. doi:10.1186/s12935-020-01413-6

Wang, Y., Yao, Y., Zhao, J., Cai, C., Hu, J., and Zhao, Y. (2021). Development of an Autophagy-Related Gene Prognostic Model and Nomogram for Estimating Renal clear Cell Carcinoma Survival. *J. Oncol.* 2021, 1–13. doi:10.1155/2021/8810849

Wu, M., Hu, Y., Ren, A., Peng, X., Ma, Q., Mao, C., et al. (2021). Nomogram Based on Ultrasonography and Clinical Features for Predicting Malignancy in Soft Tissue Tumors. *Cmar* 13, 2143–2152. doi:10.2147/CMAR.S296972

Yang, J. D., and Heimbach, J. K. (2020). New Advances in the Diagnosis and Management of Hepatocellular Carcinoma. *BMJ* 371, m3544. doi:10.1136/bmj.m3544

Zhang, Y., Zhang, D., Lv, J., Wang, S., and Zhang, Q. (2019). LncRNA SNHG15 Acts as an Oncogene in Prostate Cancer by Regulating mir-338-3p/FKBP1A axis. *Gene* 705, 44–50. doi:10.1016/j.gene.2019.04.033

Zhou, D., Liu, X., Wang, X., Yan, F., Wang, P., Yan, H., et al. (2021). A Prognostic Nomogram Based on Lasso Cox Regression in Patients with Alpha-Fetoprotein-Negative Hepatocellular Carcinoma Following Non-surgical Therapy. *BMC Cancer* 21 (1), 246. doi:10.1186/s12885-021-07916-3

Zhou, Y., Zhou, X., Ma, J., Zhang, W., Yan, Z., and Luo, J. (2021). Nomogram for Predicting the Prognosis of Patients with Hepatocellular Carcinoma Presenting with Pulmonary Metastasis. *Cmar* 13, 2083–2094. doi:10.2147/CMAR.S296020

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 December 2021
doi: 10.3389/fgene.2021.782699

ECM–Receptor Regulatory Network and Its Prognostic Role in Colorectal Cancer

Stepan Nersisyan[1*†], *Victor Novosad*[1,2†], *Narek Engibaryan*[1], *Yuri Ushkaryov*[1,2,3], *Sergey Nikulin*[1,4,5] *and Alexander Tonevitsky*[1,2,6]

[1]Faculty of Biology and Biotechnology, HSE University, Moscow, Russia, [2]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia, [3]Medway School of Pharmacy, University of Kent, Chatham, United Kingdom, [4]P. Hertsen Moscow Oncology Research Institute—Branch, National Medical Research Radiological Centre, Ministry of Health of Russian Federation, Moscow, Russia, [5]School of Biomedicine, Far Eastern Federal University, Vladivostok, Russia, [6]SRC Bioclinicum, Moscow, Russia

OPEN ACCESS

Edited by:
Yuriy L. Orlov,
I.M.Sechenov First Moscow State Medical University, Russia

Reviewed by:
Kerim Mutig,
Charité University Medicine Berlin, Germany
Arsen Arakelyan,
Institute of Molecular Biology (IMB), Armenia
Emanuela Bostjancic,
University of Ljubljana, Slovenia

***Correspondence:**
Stepan Nersisyan
snersisyan@hse.ru

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 24 September 2021
Accepted: 05 November 2021
Published: 06 December 2021

Citation:
Nersisyan S, Novosad V, Engibaryan N, Ushkaryov Y, Nikulin S and Tonevitsky A (2021) ECM–Receptor Regulatory Network and Its Prognostic Role in Colorectal Cancer. Front. Genet. 12:782699. doi: 10.3389/fgene.2021.782699

Interactions of the extracellular matrix (ECM) and cellular receptors constitute one of the crucial pathways involved in colorectal cancer progression and metastasis. With the use of bioinformatics analysis, we comprehensively evaluated the prognostic information concentrated in the genes from this pathway. First, we constructed a ECM–receptor regulatory network by integrating the transcription factor (TF) and 5'-isomiR interaction databases with mRNA/miRNA-seq data from The Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD). Notably, one-third of interactions mediated by 5'-isomiRs was represented by noncanonical isomiRs (isomiRs, whose 5'-end sequence did not match with the canonical miRBase version). Then, exhaustive search-based feature selection was used to fit prognostic signatures composed of nodes from the network for overall survival prediction. Two reliable prognostic signatures were identified and validated on the independent The Cancer Genome Atlas Rectum Adenocarcinoma (TCGA-READ) cohort. The first signature was made up by six genes, directly involved in ECM–receptor interaction: AGRN, DAG1, FN1, ITGA5, THBS3, and TNC (concordance index 0.61, logrank test p = 0.0164, 3-years ROC AUC = 0.68). The second hybrid signature was composed of three regulators: hsa-miR-32-5p, NR1H2, and SNAI1 (concordance index 0.64, logrank test p = 0.0229, 3-years ROC AUC = 0.71). While hsa-miR-32-5p exclusively regulated ECM-related genes (COL1A2 and ITGA5), NR1H2 and SNAI1 also targeted other pathways (adhesion, cell cycle, and cell division). Concordant distributions of the respective risk scores across four stages of colorectal cancer and adjacent normal mucosa additionally confirmed reliability of the models.

Keywords: ECM–receptor, colorectal cancer, network analysis, TCGA-COAD, TCGA-READ

1 INTRODUCTION

The extracellular matrix (ECM) is a noncellular component of tissue, which biochemically and structurally supports cells. The ECM is composed of different glycoproteins such as collagens, laminins, and fibronectins (Theocharis et al., 2016), and there are dozens of cellular receptors which directly interact with the components of the ECM, for example, integrins or cadherins (Barczyk et al., 2010). Interactions between the ECM and receptors on the cellular surface regulate cell behavior and

play an important role in communications between cells, cell proliferation, adhesion, and migration (Nguyen-Ngoc et al., 2012; Plotnikov et al., 2012; Schlie-Wolter et al., 2013; Lange et al., 2014).

A number of studies revealed a crucial role of the ECM–receptor interaction in colorectal cancer development and metastasis formation (Stankevicius et al., 2016; Crotti et al., 2017; Maltseva and Rodin, 2018). We recently showed the contribution of α5 laminin in differentiation of colorectal cancer cells and chemotherapy resistance (Maltseva et al., 2020). Several works describe biomarkers and signatures for assessment of colorectal cancer prognosis based on the expression of particular genes involved in ECM–receptor interaction, including genes encoding integrins (Boudjadi et al., 2013; Gong et al., 2019), E- and P-cadherin (Sun et al., 2011; Christou et al., 2017), and different laminins (Galatenko et al., 2018). ECM-based prognostic gene signatures were constructed for gastric (Yang et al., 2020), breast (Bergamaschi et al., 2008), prostate (Pang et al., 2019), and bladder (Qing et al., 2020) cancers. However, to the best of our knowledge, no comprehensive prognostic analysis of ECM–receptor interaction–based colorectal cancer gene signatures has been done so far.

Another dimension useful for the construction of prognostic signatures is the analysis of regulatory networks (Ahmad et al., 2012; Guo et al., 2020; Nersisyan et al., 2021b). Specifically, gene expression levels could be dynamically regulated by other molecules, such as transcription factors (TFs), microRNAs (miRNAs), and others. Recently, it was shown that miRNAs are present in a cell in different variants, called miRNA isoforms (isomiRs) (Morin et al., 2008; Loher et al., 2014). As a result of imprecise enzymatic cleavage, miRNA hairpins give rise to mature forms, which differ from each other in 1–3 nucleotides at the ends of the molecule (Zhiyanov et al., 2021). Importantly, the targetome of isomiRs with differences at 5'-ends (5'-isomiRs) significantly differ from the canonical form (Tan et al., 2014; van der Kwast et al., 2020). Thus, 5'-isomiRs could be considered separate miRNAs with their own sets of targets.

In this work, we analyzed expression patterns of genes involved in the ECM–receptor interaction pathway using RNA sequencing data of colorectal cancer samples taken from The Cancer Genome Atlas Colon Adenocarcinoma (TCGA-COAD) and Rectum Adenocarcinoma (TCGA-READ) projects (Network, 2012). First, we constructed and analyzed a regulatory network to infer 5'-isomiRs and TFs, which are direct regulators of genes from the ECM–receptor interaction pathway. The network was built with miRGTF-net—the recently developed tool which integrates both expression and database-level data for the network construction (Nersisyan et al., 2021b). Next, the obtained network was used to construct hybrid isomiR-gene signatures for predicting overall survival in colorectal cancer. For this analysis, we employed a novel technique of exhaustive search-based Cox model fitting. Namely, ExhauFS software (Nersisyan et al., 2021c) was used to construct prognostic models for all gene/5'-isomiR pairs, triples, etc; then, the best performing model was picked.

2 METHODS

2.1 TCGA mRNA and miRNA Sequencing Data

RNA and miRNA sequencing read count tables were downloaded from the GDC Data Portal for $n = 426$ TCGA-COAD and $n = 161$ TCGA-READ colorectal cancer samples (tumors with unmatched miRNA/mRNA profiles or without clinical information were not considered). For the comparison of primary tumors and adjacent normal mucosa, $n = 7$ TCGA-COAD normal samples were also included. With the use of the trimmed mean of M-values (TMM) algorithm implemented in the edgeR v3.30.3 package (Robinson et al., 2010), the obtained mRNA-seq and miRNA-seq count matrices were processed into the TMM-FPKM and TMM-RPM tables, respectively. Low expressed genes and miRNAs were filtered out using the default procedure available in edgeR.

The conventional nomenclature was used to annotate 5'-isomiRs (Telonis et al., 2015; Zhiyanov et al., 2021). For example, hsa-miR-30e-5p|+1 stands for the mature hsa-miR-30e-5p miRNA without the first nucleotide at the 5'-end (i.e., the number after | represents the offset at the 5'-end in the direction from the 5'-end to the 3'-end).

2.2 Network Analysis

A recently developed miRGTF-net tool (Nersisyan et al., 2021b) was applied to the TCGA-COAD dataset for the construction of a colorectal cancer miRNA-gene-TF regulatory network. The main feature of this approach consists in the integration of expression data (TCGA-COAD) with the biological interaction databases:

- TFLink database (https://tflink.net) was used to extract TF-gene interactions;
- TF-miRNA interactions were obtained from TransmiR v2.0 (Tong et al., 2019);
- miRDB v6.0 (Chen and Wang, 2020) custom prediction mode was employed to predict targets of 5′-isomiRs (as recommended by the tool authors, interactions with target scores ≥ 80 were considered).

First, the initial network was constructed based on the interactions listed in these databases. Second, all uncorrelated and wrong-directional edges (like positively correlated miRNAs and their targets) were discarded. Then, interaction scores were assigned to each edge and node of the network. The interaction scores are based on the strength of the linear dependence between expressions levels of the connected nodes. After filtering out nodes and edges with low interaction scores, the resulting network consisted of nodes with significant influence on some other nodes and/or significantly regulated by some other node. Configuration for miRGTF-net execution is listed in **Supplementary Material**.

ECM–receptor interaction–related genes were taken from the KEGG hsa04512 pathway (Kanehisa et al., 2021) (we refer to these genes as *ECM set*). Next, the output of miRGTF-net was used to construct a subnetwork composed of molecules, which

either regulates a gene from the ECM–receptor interaction pathway or is regulated by a gene from the pathway (*ECM+ set*).

2.3 Construction of Prognostic Signatures

The TCGA-COAD cohort was split into training (75% of samples) and filtration (25%) sets with a stratification by outcome indicator (death or censoring) and overall survival time (date of death or date of the last follow-up). Namely, we first sorted samples by outcome indicator and then sorted samples by overall survival time within each outcome group. Finally, every fourth sample was added to the filtration set, and the resting samples were labeled as the training ones. The independent TCGA-READ dataset was used for the model validation (test set).

The ExhauFS tool (Nersisyan et al., 2021c) was used to fit Cox survival regression models. For each length of prognostic signature ($k = 1, 2, \ldots, 10$), we first selected n most individually predictive features (see the next paragraph for the details) and then fit Cox models for all possible $\binom{n}{k}$ feature subsets. The values of n were chosen for reasons of limiting the computational time by the default procedure available in ExhauFS (**Supplementary Table S1**). The pipeline was executed in two modes: in the first one, genes from the ECM set were pre-selected, and in the second run, all 537 genes and isomiRs from the ECM+ set were considered for the model construction.

The concordance index was used as the main model accuracy metric, including the feature selection step. That is, features (genes and isomiRs) were selected according to the concordance index of the respective univariate model. In addition, patients were separated into high- and low-risk groups (the median risk score calculated on the training set was used as a cut-off value). This allowed us to construct Kaplan–Meier curves and compare low- and high-risk groups with the hazard ratio metric and the logrank test. Finally, time-dependent ROC AUC was calculated to measure discriminative power of models for predicting 3-year patient survival. Configuration for ExhauFS execution is listed in **Supplementary Material**.

The set of reliable models was defined by the following thresholds, set on both training and filtration sets: concordance index > 0.6, hazard ratio > 2, logrank test p-value < 0.01, and 3-year ROC AUC > 0.6. The best performing model was chosen by taking the signature with the maximal concordance index on the training set.

2.4 Enrichment Analysis

Enrichment analysis of gene sets was conducted using DAVID v6.8 (Huang et al., 2009). Significantly enriched terms were identified by setting a 0.05 threshold on false discovery rates (FDRs).

2.5 Statistical Analysis

A hypergeometric test was applied to

- identify regulators (TFs and isomiRs) with an overrepresented number of target genes in the ECM set;
- identify genes and isomiRs, which were overrepresented in the reliable prognostic signatures.

"Over"-regulated genes from the ECM set were determined by the binomial test. In all the cases, the Benjamini–Hochberg procedure was employed to adjust for multiple testing correction. SciPy implementation of statistical tests was used (Virtanen et al., 2020).

3 RESULTS

3.1 Regulatory Network of ECM–Receptor Interaction Pathway

The first step of our analysis was the inference of regulatory interactions affecting genes from the ECM–receptor interaction pathway (from here onward, we refer to these genes as *ECM set*). The MiRGTF-net tool allows one to construct miRNA–gene–TF interaction networks combining both database-level and integrative miRNA/gene expression data. At the beginning, the database-level network was constructed; it contained interactions of the three types:

- TFs regulating target genes;
- TFs regulating target miRNAs;
- 5'-isomiRs downregulating target genes.

Then, TCGA-COAD gene and isomiR expression data were analyzed to filter only those interactions which are supported by a significant correlation in considered samples.

The resulting ECM–receptor regulatory network consisted of 522 nodes, which included 442 TFs, 27 5'-isomiRs, and 53 genes from the ECM–receptor interaction pathway (here onward, *ECM+ set*). First, we analyzed out-degrees of the network nodes, that is, the numbers of regulatory interactions outgoing from TFs and isomiRs. The network had 49 hubs—regulators whose targets were significantly enriched by the ECM set (**Figure 1A**, **Supplementary Table S2**). Aside from well-known TFs, regulating hundreds and thousands of genes (e.g., ZEB1, TWIST1, SPI1, etc.), there were three isomiRs (hsa-miR-148a-3p|-1, hsa-miR-29b-3p|0, and hsa-miR-32–5p|0) narrowly regulating the ECM set. Reciprocally, network in-degree analysis revealed nine "over"-regulated genes from the ECM set, mainly integrins and laminins (**Figure 1B**, **Supplementary Table S3**).

Interestingly, one-third of isomiR–target gene interactions (13 out of 39) were mediated by noncanonical 5'-isomiRs (i.e., isomiRs whose 5'-ends do not match with canonical the miRBase version). This included four mRNA targets of hsa-miR-148a-3p|-1 (LAMA4, LAMB2, ITGA11, and COL4A1), two targets of hsa-miR-335-3p|-1 and hsa-miR-30e-5p|+1 (both isomiRs regulated ITGA1 and COL1A2), and five isolated isomiR–gene interactions: hsa-miR-92a-3p|+2 and ITGAV, hsa-miR-203a-3p|+1 and COL4A1, hsa-miR-200b-3p|+1 and LAMA4, hsa-miR-194-5p|-1 and ITGA2, and hsa-miR-142-3p|+1 and LAMC1. Thus, consideration of 5'-isomiRs as distinct functional units

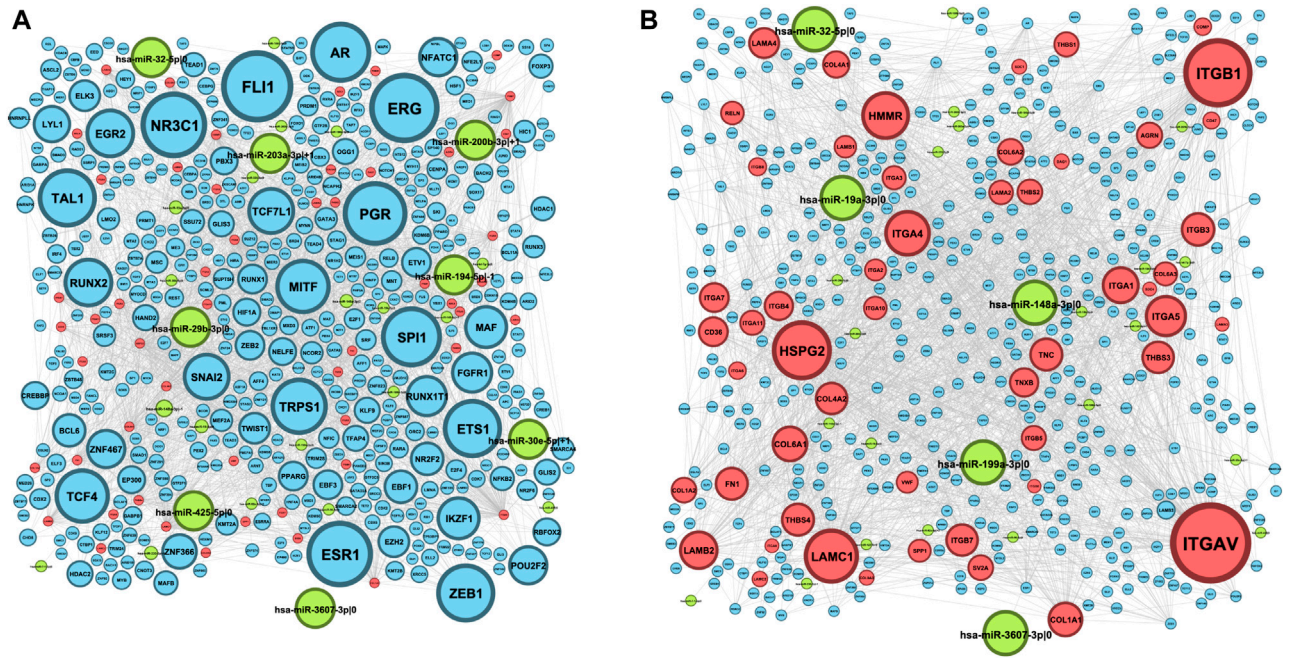


FIGURE 1 | ECM–receptor regulatory network. Blue nodes represent TFs, red nodes represent genes, and green nodes represent 5′-isomiRs. The node size indicates the value of the node degree; **(A)** out-degree; **(B)** in-degree.

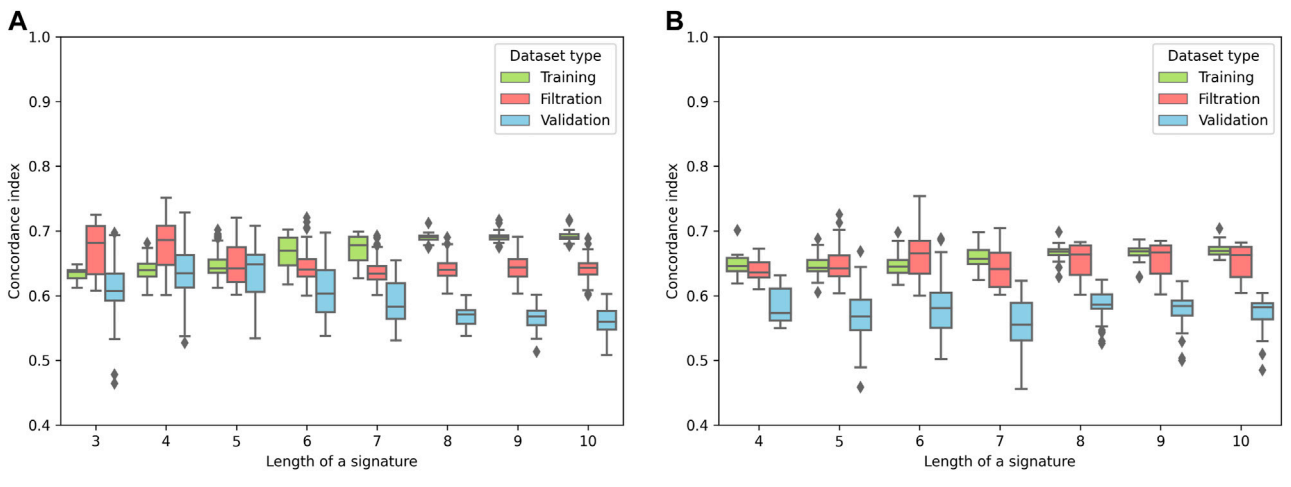


FIGURE 2 | Distribution of the concordance indices of the reliable models for different lengths of signatures. **(A)** ECM+ models. **(B)** ECM models.

added much information about RNA interference–mediated gene silencing.

3.2 Prognostic Power of the ECM and ECM+ Sets

Next, we assessed whether it is possible to find an accurate overall survival prediction model constructed of molecules from the ECM and ECM+ sets. We used our recently developed ExhauFS tool to go over all possible prognostic signatures composed of ECM/ECM+ genes and 5'-isomiRs, where the signature length varied from 1 to 10. While it was possible to search over all possible gene/isomiR pairs composed of 537 molecules (cardinality of the ECM+ set), the exhaustive search was computationally infeasible already for the triples. To tackle this problem, ExhauFS selects the relevant number of the most individually informative features and then performs exhaustive search among them (see Methods for the details).

For the pipeline evaluation, 75% of the TCGA-COAD cohort was used for the Cox model training, and the remaining 25% was

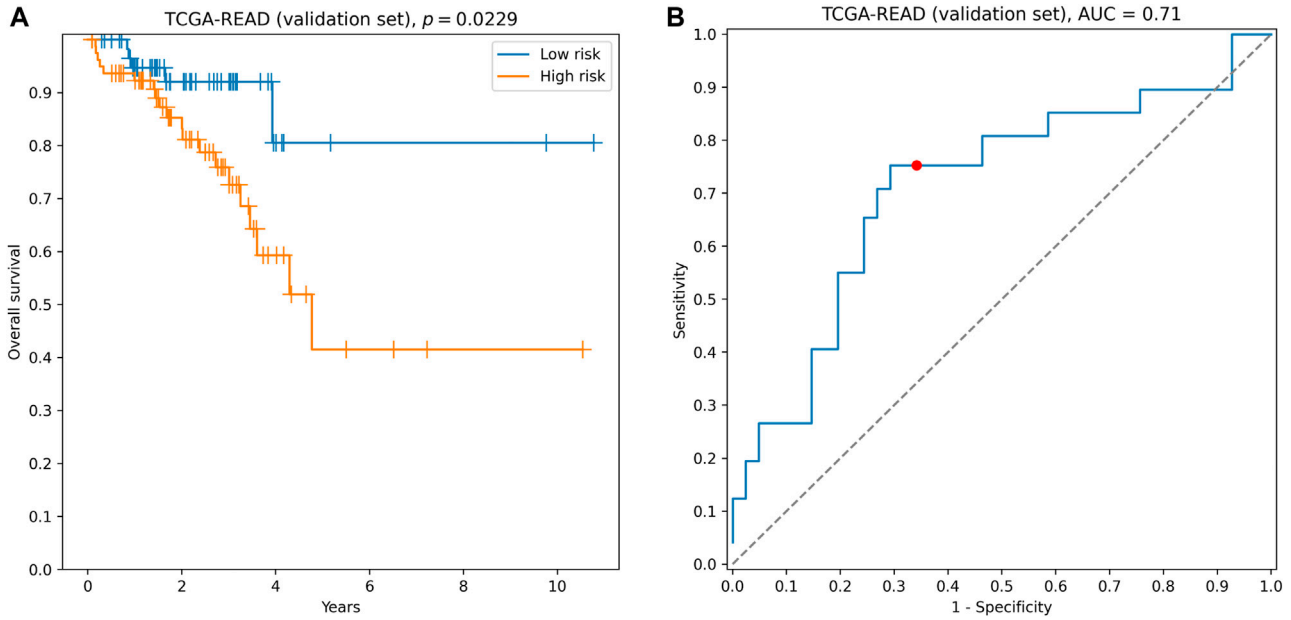


FIGURE 3 | Performance of hsa-miR-32-5p|0, NR1H2, and SNAI1 signature on the validation set. **(A)** Kaplan–Meier curves. **(B)** 3-year ROC curve. Red point on the ROC curve corresponds to the risk score threshold, calculated as a median score on the training set.

used for the filtration. The TCGA-READ dataset was used as an independent validation set. We set up several accuracy thresholds to discard models which demonstrated unreliable quality either on the training or the filtration sets (see Methods). The distribution of model accuracies (concordance indices) for each signature length (k) is shown in **Figure 2**. As it can be seen, both ECM+ and ECM models started to overfit from some point: quality of the models monotonically increased on the training set and started to drop on the filtration set after a certain signature length. The filtration set accuracy peak for the ECM+ set fell on gene/isomiR triples and quadruples (**Figure 2**); for the downstream analysis, we selected the shorter signatures, since there was no statistically significant difference between concordance indices for $k = 3$ and $k = 4$ (Mann–Whitney U-test $p = 0.11$). As for the ECM set, the highest filtration concordance indices were detected for the 6-gene signatures.

Among the ECM+ triples, the best model (according to the concordance index on the training set) was constructed with one canonical miRNA and two TFs: hsa-miR-32-5p|0, NR1H2, and SNAI1. The risk score (RS) for the model was calculated as follows:

$$RS = 0.25 * \text{hsa-miR-32-5p|0} + 0.34 * \text{NR1H2} + 0.25 * \text{SNAI1}.$$

The signature demonstrated reliable performance on the TCGA-READ validation set: the concordance index was equal to 0.64, difference in survival between groups of low- and high-risk was statistically significant (hazard ratio = 2.25, logrank test $p = 0.0229$, **Figure 3A**), and the model accurately classified 3-year patient survival (3-year ROC AUC = 0.71, **Figure 3B**).

Similarly to the ECM+ case, the most reliable signature in the ECM set was identified as follows:

$$RS = 0.30 * \text{AGRN} - 0.57 * \text{DAG1} + 0.11 * \text{FN1} + 0.36 * \text{ITGA5} + 0.29 * \text{THBS3} - 0.42 * \text{TNC}.$$

The quality of this signature, composed of six genes directly involved in ECM–receptor interaction, was comparable to the quality of hybrid ECM+ prognostic triple: concordance index = 0.61, hazard ratio = 2.14, logrank test $p = 0.0164$ (**Figure 4A**), 3-year ROC AUC = 0.68 (**Figure 4B**). The complete list of accuracy metrics (including training and filtration sets) is presented in **Supplementary Table S4** and **Supplementary Figure S1**.

To assess the relationship between expression levels of nine identified prognostic molecules, we performed hierarchical clustering using both sample-wise expression values (**Figure 5A**) and correlation matrix (**Figure 5B**). Notably, three genes (FN1, ITGA5, and TNC) showed a strong co-expression pattern, while the other molecules did not form clear cluster structures. For both signatures, we compared the distributions of the underlying risk scores between four stages of colorectal cancer and adjacent normal tissues. In all cases, the risk scores monotonically increased from the normal mucosa to stage IV cancer (**Figure 6**). This observation is an additional piece of evidence of reliability of two constructed models.

3.3 Regulatory Neighborhood of the Prognostic 5'-isomiR/Gene Triple

Since the best ECM+ model was composed of three regulators (one miRNA and two TFs), the next step of our analysis was to

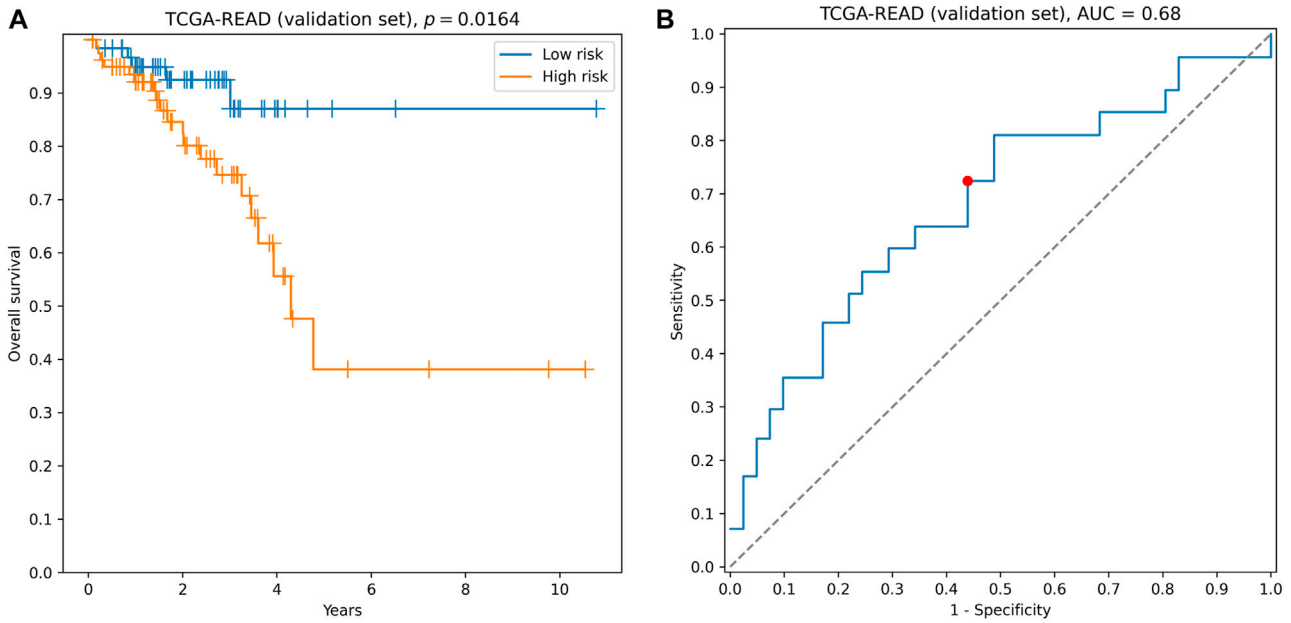


FIGURE 4 | Performance of AGRN, DAG1, FN1, ITGA5, THBS3, and TNC signature on the validation set. **(A)** Kaplan–Meier curves. **(B)** 3-year ROC curve. Red point on the ROC curve corresponds to the risk score threshold, calculated as a median score on the training set.

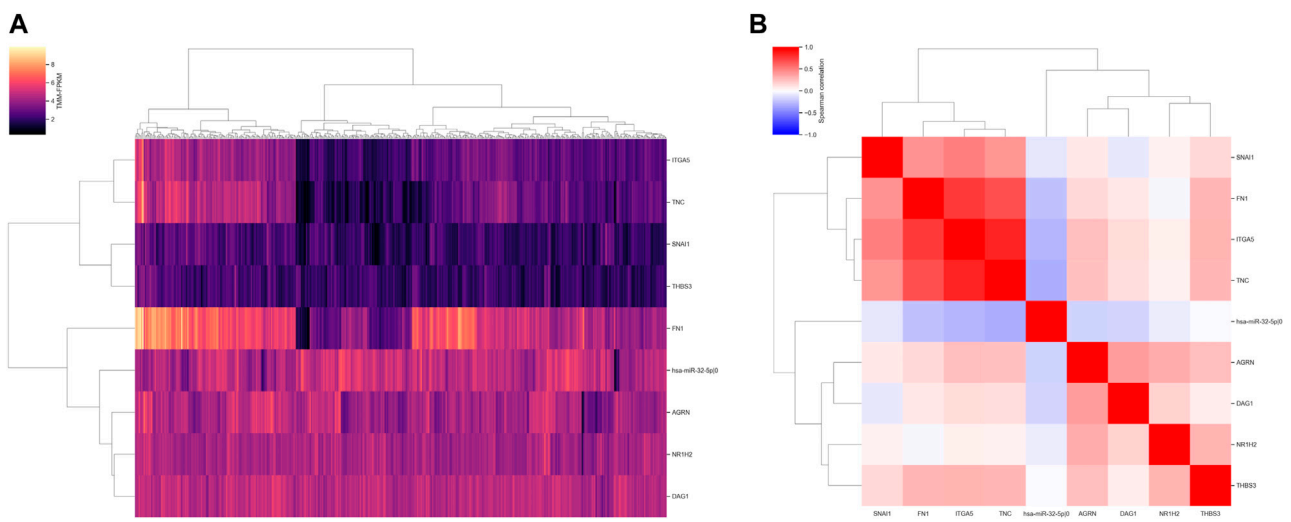


FIGURE 5 | Expression distribution of the nine prognostic molecules. **(A)** Normalized and $\log_2$-transformed expression units. **(B)** Spearman correlation matrix.

explore the landscape of regulatory interactions mediated by these molecules (**Figure 7**). Out of three regulators, only miR-32 specifically regulated the ECM–interaction pathway: two (COL1A2 and ITGA5) out of nine predicted targets were from the ECM set (adjusted $p = 8.92 \times 10^{-3}$). Moreover, DAVID enrichment analysis of this set of nine genes revealed only two pathways tightly related to the ECM: ECM–receptor interaction (KEGG hsa04512, adjusted $p = 0.0168$) and focal adhesion (KEGG hsa04510, $p = 0.0463$).

Unlike miR-32, targetomes of NR1H2 and SNAI1 were not focused on the ECM set; only 4/705 targets of NR1H2 (HMMR, ITGA3, ITGB4, and LAMB3, $p = 0.86$) and 1/6 targets of SNAI1 (FN1, $p = 0.16$) were associated with ECM–receptor interaction. To uncover the regulatory role of NR1H2 and SNAI1 in colorectal cancer, we also performed DAVID functional enrichment analysis of the sets of their target genes inferred by miRGTF-net. Multiple pathways were enriched in the set of NR1H2 targets, including cell cycle, cell division, DNA repair, and RNA splicing (**SupplementaryTable S5**). In case of SNAI1, the cell adhesion pathway was enriched when no multiple testing correction was applied (**Supplementary Table S5**). Thus, the inclusion of

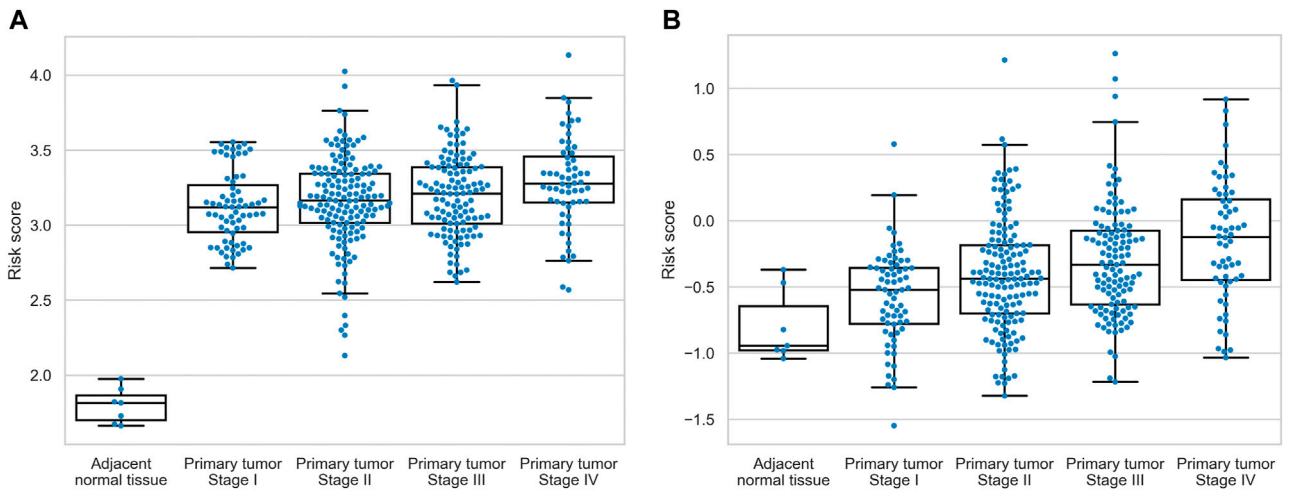


FIGURE 6 | Distribution of the risk scores across normal tissues and four stages of colorectal cancer. **(A)** hsa-miR-32-5p|0, NR1H2, and SNAI1 signature. **(B)** AGRN, DAG1, FN1, ITGA5, THBS3, and TNC signature.

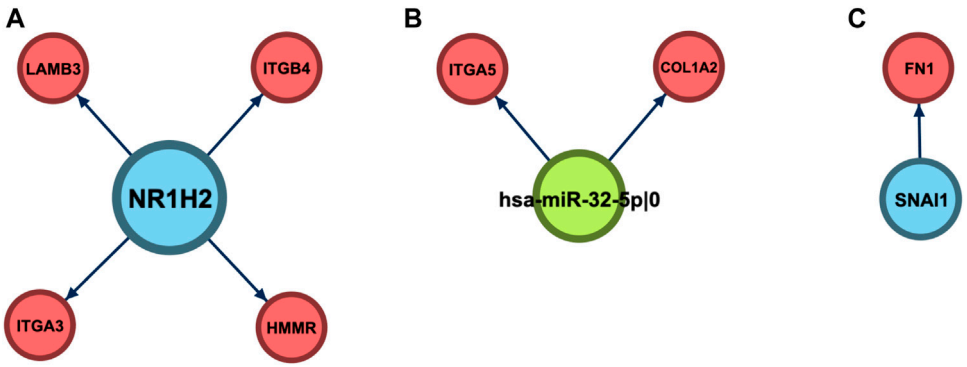


FIGURE 7 | Regulatory neighborhood of the prognostic 5′-isomiR/gene triple. **(A)** NR1H2. **(B)** hsa-miR-32–5p|0. **(C)** SNAI1.

regulators in the prognostic signatures expanded the scope of the considered ECM pathway.

4 DISCUSSION

In this work, we used expression data of genes constituting the ECM–receptor interaction pathway and its direct 5'-isomiR and TF regulators to compose prognostic signatures for colorectal cancer. The novel feature of the network construction step consisted in accounting for 5'-isomiR targeting. Importantly, one-third of all isomiR–gene interactions were mediated by noncanonical 5'-isomiRs. These numbers are in agreement with previous experimental findings, which demonstrated biological activity of noncanonical 5'-isomiRs, for example, miR-411|-1 (van der Kwast et al., 2020) or miR-9|+1 (Tan et al., 2014).

With the use of the ECM–receptor regulatory network, we constructed two reliable signatures for overall survival prediction. The first hybrid signature was composed of one canonical miRNA and two TFs: hsa-miR-32-5p|0, NR1H2, and SNAI1. The second signature was composed of six genes, directly involved in ECM–receptor interaction: AGRN, DAG1, FN1, ITGA5, THBS3, and TNC. A number of studies already highlighted the role of several markers from the constructed signatures in colorectal cancer. In two recent reports, miR-32 was shown to promote tumorigenesis, radioresistance, migration, and invasion of colorectal cancer by targeting BMP5 and TOB1 (Chen et al., 2018; Liang et al., 2019). With the use of sequence-based target prediction coupled with co-expression analysis, here, we first showed that miR-32 targets are overrepresented in the ECM–receptor interaction pathway (adjusted p-value = 8.92×10^{-3}). Thus, the new possible regulatory role of miR-32 was uncovered. Another member of the ECM+ prognostic triple, the SNAI1 transcription factor, was also linked to the poor prognosis of colorectal cancer. Namely, SNAI1 regulates epithelial–mesenchymal transition (EMT) by suppressing E-cadherin and promotes chemoresistance in colorectal cancer (Hoshino et al., 2009; Wang et al., 2018). Nevertheless, we did not find specific roles of NR1H2 in the colorectal cancer development,

progression, or metastasis. Functional enrichment analysis of NR1H2 downstream targets suggested the contribution of this TF to the regulation of core cellular pathways, such as cell cycle and cell division.

The second signature was also partially composed of well-studied genes. We previously showed the strong upregulation of ITGA5 in Caco-2 human colorectal cancer cell lines exposed to hypoxia, as a consequence of hypoxia-induced decrease in expression of its direct regulator—miR-148a (Nersisyan et al., 2021a). Notably, this regulatory interaction (miR-148a suppressing ITGA5) was also supported by the negative correlation in our miRGTF-net analysis. In the same work, the negative association between ITGA5 expression levels and patients' overall survival was observed. Other studies showed that reduced DAG1 protein expression is associated with poor outcome of colorectal cancer (Coco et al., 2012); downregulation of FN1 decreases proliferation, migration, and invasion of colorectal cancer cells (Cai et al., 2018), while TNC induces EMT and proliferation (Yang et al., 2018). As for AGRN and THBS3, we have not found evidence on their role in colorectal cancer pathogenesis. The comprehensive reference list summarizing the role of the selected genes in colorectal cancer prognosis is presented in **Supplementary Table S6**.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Conceptualization: SNe, YU, and AT. Methodology: SNe, YU, and AT. Data curation: SNe. Execution of bioinformatics software: SNi. Vizualization: VN and SNi. Investigation: SNe, VN, and AT. Writing—original draft: SNe, VN, NE, and SNi. Writing—review and editing: SNe, VN, NE, YU, SNi, and AT.

FUNDING

The research was performed within the framework of the Basic Research Program at HSE University (SNe, YU, and AT; conceptualization, methodology, data curation, and investigation) and the Russian Science Foundation (Project No. 19-15-00397; SNi; bioinformatics analysis and vizualization).

ACKNOWLEDGMENTS

The authors thank Alexey Makhonin for the composition of the isomiR–gene interaction list.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.782699/full#supplementary-material

REFERENCES

Ahmad, F. K., Deris, S., and Othman, N. H. (2012). The Inference of Breast Cancer Metastasis through Gene Regulatory Networks. *J. Biomed. Inform.* 45, 350–362. doi:10.1016/j.jbi.2011.11.015

Barczyk, M., Carracedo, S., and Gullberg, D. (2010). Integrins. *Cell Tissue Res* 339, 269–280. doi:10.1007/s00441-009-0834-6

Bergamaschi, A., Tagliabue, E., Sørlie, T., Naume, B., Triulzi, T., Orlandi, R., et al. (2008). Extracellular Matrix Signature Identifies Breast Cancer Subgroups with Different Clinical Outcome. *J. Pathol.* 214, 357–367. doi:10.1002/path.2278

Boudjadi, S., Carrier, J. C., and Beaulieu, J.-F. (2013). Integrin α1 Subunit Is Up-Regulated in Colorectal Cancer. *Biomark Res.* 1, 16. doi:10.1186/2050-7771-1-16

Cai, X., Liu, C., Zhang, T. N., Zhu, Y. W., Dong, X., and Xue, P. (2018). Down-regulation of FN1 Inhibits Colorectal Carcinogenesis by Suppressing Proliferation, Migration, and Invasion. *J. Cel. Biochem.* 119, 4717–4728. doi:10.1002/jcb.26651

Chen, E., Li, Q., Wang, H., Zhang, P., Zhao, X., Yang, F., et al. (2018). MiR-32 Promotes Tumorigenesis of Colorectal Cancer by Targeting BMP5. *Biomed. Pharmacother.* 106, 1046–1051. doi:10.1016/j.biopha.2018.07.050

Chen, Y., and Wang, X. (2020). miRDB: an Online Database for Prediction of Functional microRNA Targets. *Nucleic Acids Res.* 48, D127–D131. doi:10.1093/nar/gkz757

Christou, N., Perraud, A., Blondy, S., Jauberteau, M.-O., Battu, S., and Mathonnet, M. (2017). E-cadherin: A Potential Biomarker of Colorectal Cancer Prognosis. *Oncol. Lett.* 13, 4571–4576. doi:10.3892/ol.2017.6063

Coco, C., Zannoni, G. F., Caredda, E., Sioletic, S., Boninsegna, A., Migaldi, M., et al. (2012). Increased Expression of CD133 and Reduced Dystroglycan Expression Are strong Predictors of Poor Outcome in colon Cancer Patients. *J. Exp. Clin. Cancer Res.* 31, 71. doi:10.1186/1756-9966-31-71

Crotti, S., Piccoli, M., Rizzolio, F., Giordano, A., Nitti, D., and Agostini, M. (2017), Extracellular Matrix and Colorectal Cancer: How Surrounding Microenvironment Affects Cancer Cell Behavior? *J. Cel. Physiol.* 232, 967–975. doi:10.1002/jcp.25658

Galatenko, V. V., Maltseva, D. V., Galatenko, A. V., Rodin, S., and Tonevitsky, A. G. (2018). Cumulative Prognostic Power of Laminin Genes in Colorectal Cancer. *BMC Med. Genomics* 11, 9. doi:10.1186/s12920-018-0332-3

Gong, Y., Ruan, G., Liao, X., Wang, X., Liao, C., Wang, S., et al. (2019). Diagnostic and Prognostic Values of Integrin *α* Subfamily mRNA Expression in colon Adenocarcinoma. *Oncol. Rep.* doi:10.3892/or.2019.7216

Guo, Q., Wang, J., Gao, Y., Li, X., Hao, Y., Ning, S., et al. (2020). Dynamic TF-lncRNA Regulatory Networks Revealed Prognostic Signatures in the Development of Ovarian Cancer. *Front. Bioeng. Biotechnol.* 8. doi:10.3389/fbioe.2020.00460

Hoshino, H., Miyoshi, N., Nagai, K.-i., Tomimaru, Y., Nagano, H., Sekimoto, M., et al. (2009). Epithelial-mesenchymal Transition with Expression of SNAI1-Induced Chemoresistance in Colorectal Cancer. *Biochem. Biophysical Res. Commun.* 390, 1061–1065. doi:10.1016/j.bbrc.2009.10.117

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4, 44–57. doi:10.1038/nprot.2008.211

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: Integrating Viruses and Cellular Organisms. *Nucleic Acids Res.* 49, D545–D551. doi:10.1093/nar/gkaa970

Lange, T., Samatov, T. R., Tonevitsky, A. G., and Schumacher, U. (2014). Importance of Altered Glycoprotein-Bound N- and O-Glycans for Epithelial-To-Mesenchymal Transition and Adhesion of Cancer Cells. *Carbohydr. Res.* 389, 39–45. doi:10.1016/j.carres.2014.01.010

Liang, H., Tang, Y., Zhang, H., and Zhang, C. (2019). MiR-32-5p Regulates Radiosensitization, Migration and Invasion of Colorectal Cancer Cells by Targeting TOB1 Gene. *Ott* Vol. 12, 9651–9661. doi:10.2147/OTT.S228995

Loher, P., Londin, E. R., and Rigoutsos, I. (2014). IsomiR Expression Profiles in Human Lymphoblastoid Cell Lines Exhibit Population and Gender Dependencies. *Oncotarget* 5, 8790–8802. doi:10.18632/oncotarget.2405

Maltseva, D., Raygorodskaya, M., Knyazev, E., Zgoda, V., Tikhonova, O., Zaidi, S., et al. (2020). Knockdown of the α5 Laminin Chain Affects Differentiation of Colorectal Cancer Cells and Their Sensitivity to Chemotherapy. *Biochimie* 174, 107–116. doi:10.1016/j.biochi.2020.04.016

Maltseva, D. V., and Rodin, S. A. (2018). Laminins in Metastatic Cancer. *Mol. Biol.* 52, 350–371. doi:10.1134/s0026893318030093

Morin, R. D., O'Connor, M. D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.-L., et al. (2008). Application of Massively Parallel Sequencing to microRNA Profiling and Discovery in Human Embryonic Stem Cells. *Genome Res.* 18, 610–621. doi:10.1101/gr.7179508

Nersisyan, S., Galatenko, A., Chekova, M., and Tonevitsky, A. (2021a). Hypoxia-Induced miR-148a Downregulation Contributes to Poor Survival in Colorectal Cancer. *Front. Genet.* 12. doi:10.3389/fgene.2021.662468

Nersisyan, S., Galatenko, A., Galatenko, V., Shkurnikov, M., and Tonevitsky, A. (2021b). miRGTF-Net: Integrative miRNA-Gene-TF Network Analysis Reveals Key Drivers of Breast Cancer Recurrence. *PLOS ONE* 16, e0249424. doi:10.1371/journal.pone.0249424

Nersisyan, S., Novosad, V., Galatenko, A., Sokolov, A., Bokov, G., Konovalov, A., et al. (2021c). Exhaufs: Exhaustive Search-Based Feature Selection for Classification and Survival Regression. *bioRxiv*. doi:10.1101/2021.08.03.454798

Network, T. C. G. A. (2012). Comprehensive Molecular Characterization of Human colon and Rectal Cancer. *Nature* 487, 330–337. doi:10.1038/nature11252

Nguyen-Ngoc, K.-V., Cheung, K. J., Brenot, A., Shamir, E. R., Gray, R. S., Hines, W. C., et al. (2012). ECM Microenvironment Regulates Collective Migration and Local Dissemination in normal and Malignant Mammary Epithelium. *Proc. Natl. Acad. Sci.* 109, E2595–E2604. doi:10.1073/pnas.1212834109

Pang, X., Xie, R., Zhang, Z., Liu, Q., Wu, S., and Cui, Y. (2019). Identification of SPP1 as an Extracellular Matrix Signature for Metastatic Castration-Resistant Prostate Cancer. *Front. Oncol.* 9, 924. doi:10.3389/fonc.2019.00924

Plotnikov, S. V., Pasapera, A. M., Sabass, B., and Waterman, C. M. (2012). Force Fluctuations within Focal Adhesions Mediate ECM-Rigidity Sensing to Guide Directed Cell Migration. *Cell* 151, 1513–1527. doi:10.1016/j.cell.2012.11.034

Qing, L., Gu, P., Liu, M., Shen, J., Liu, X., Guang, R., et al. (2020). Extracellular Matrix-Related Six-lncRNA Signature as a Novel Prognostic Biomarker for Bladder Cancer. *Ott* Vol. 13, 12521–12538. doi:10.2147/ott.S284167

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* 26, 139–140. doi:10.1093/bioinformatics/btp616

Schlie-Wolter, S., Ngezahayo, A., and Chichkov, B. N. (2013). The Selective Role of ECM Components on Cell Adhesion, Morphology, Proliferation and Communication *In Vitro*. *Exp. Cel Res.* 319, 1553–1561. doi:10.1016/j.yexcr.2013.03.016

Stankevicius, V., Vasauskas, G., Noreikiene, R., Kuodyte, K., Valius, M., and Suziedelis, K. (2016). Extracellular Matrix-dependent Pathways in Colorectal Cancer Cell Lines Reveal Potential Targets for Anticancer Therapies. *Ar* 36, 4559–4568. doi:10.21873/anticanres.11004

Sun, L., Hu, H., Peng, L., Zhou, Z., Zhao, X., Pan, J., et al. (2011). P-cadherin Promotes Liver Metastasis and Is Associated with Poor Prognosis in Colon Cancer. *Am. J. Pathol.* 179, 380–390. doi:10.1016/j.ajpath.2011.03.046

Tan, G. C., Chan, E., Molnar, A., Sarkar, R., Alexieva, D., Isa, I. M., et al. (2014). 5′ isomiR Variation Is of Functional and Evolutionary Importance. *Nucleic Acids Res.* 42, 9424–9435. doi:10.1093/nar/gku656

Telonis, A. G., Loher, P., Jing, Y., Londin, E., and Rigoutsos, I. (2015). Beyond the One-Locus-One-miRNA Paradigm: microRNA Isoforms Enable Deeper Insights into Breast Cancer Heterogeneity. *Nucleic Acids Res.* 43, 9158–9175. doi:10.1093/nar/gkv922

Theocharis, A. D., Skandalis, S. S., Gialeli, C., and Karamanos, N. K. (2016). Extracellular Matrix Structure. *Adv. Drug Deliv. Rev.* 97, 4–27. doi:10.1016/j.addr.2015.11.001

Tong, Z., Cui, Q., Wang, J., and Zhou, Y. (2019). TransmiR v2.0: an Updated Transcription Factor-microRNA Regulation Database. *Nucleic Acids Res.* 47, D253–D258. doi:10.1093/nar/gky1023

van der Kwast, R. V. C. T., Woudenberg, T., Quax, P. H. A., and Nossent, A. Y. (2020). MicroRNA-411 and its 5′-IsomiR Have Distinct Targets and Functions and Are Differentially Regulated in the Vasculature under Ischemia. *Mol. Ther.* 28, 157–170. doi:10.1016/j.ymthe.2019.10.002

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* 17, 261–272. doi:10.1038/s41592-019-0686-2

Wang, S., Yan, S., Zhu, S., Zhao, Y., Yan, J., Xiao, Z., et al. (2018). FOXF1 Induces Epithelial-Mesenchymal Transition in Colorectal Cancer Metastasis by Transcriptionally Activating SNAI1. *Neoplasia* 20, 996–1007. doi:10.1016/j.neo.2018.08.004

Yang, X., Chen, L., Mao, Y., Hu, Z., and He, M. (20202020). Progressive and Prognostic Performance of an Extracellular Matrix-Receptor Interaction Signature in Gastric Cancer. *Dis. Markers* 2020, 1–23. doi:10.1155/2020/8816070

Yang, Z., Zhang, C., Qi, W., Cui, C., Cui, Y., and Xuan, Y. (2018). Tenascin-C as a Prognostic Determinant of Colorectal Cancer through Induction of Epithelial-To-Mesenchymal Transition and Proliferation. *Exp. Mol. Pathol.* 105, 216–222. doi:10.1016/j.yexmp.2018.08.009

Zhiyanov, A., Nersisyan, S., and Tonevitsky, A. (2021). Hairpin Sequence and Structure Is Associated with Features of isomiR Biogenesis. *RNA Biol.* 1, 1–9. doi:10.1080/15476286.2021.1952759

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 20 December 2021
doi: 10.3389/fgene.2021.776984

Systematic Analysis of mRNAs and ncRNAs in BMSCs of Senile Osteoporosis Patients

*Yiyun Geng[1,2], Jinfu Chen[1], Chongfei Chang[1], Yifen Zhang[1], Li Duan[1,3], Weimin Zhu[1], Lisha Mou[1], Jianyi Xiong[1,3] and Daping Wang[1,3]**

[1]Shenzhen Key Laboratory of Tissue Engineering, Shenzhen Second People's Hospital (The First Hospital Affiliated to Shenzhen University), Shenzhen, China, [2]School of Biotechnology and Food Engineering, Changshu Institute of Technology, Suzhou, China, [3]Guangdong Provincial Research Center for Artificial Intelligence and Digital Orthopedic Technology, Shenzhen, China

Senile osteoporosis (SOP) is a worldwide age-related disease characterized by the loss of bone mass and decrease in bone strength. Bone mesenchymal stem cells (BMSCs) play an important role in the pathology of senile osteoporosis. Abnormal expression and regulation of non-coding RNA (ncRNA) are involved in a variety of human diseases. In the present study, we aimed to identify differentially expressed mRNAs and ncRNAs in senile osteoporosis patient-derived BMSCs via high-throughput transcriptome sequencing in combination with bioinformatics analysis. As a result, 415 mRNAs, 30 lncRNAs, 6 circRNAs and 27 miRNAs were found to be significantly changed in the senile osteoporosis group. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were applied to analyze the function of differentially expressed mRNAs and ncRNAs. The circRNA–miRNA–mRNA regulatory network was constructed using the cytoHubba plugin based on the Cytoscape software. Interestingly, circRNA008876-miR-150-5p-mRNA was the sole predicted circRNA-miRNA-mRNA network. The differential expression profile of this ceRNA network was further verified by qRT-PCR. The biological function of this network was validated by overexpression and knockdown experiments. In conclusion, circRNA008876-miR-150-5p-mRNA could be an important ceRNA network involved in senile osteoporosis, which provides potential biomarkers and therapeutic targets for senile osteoporosis.

Keywords: senile osteoporosis, human bone marrow mesenchymal stem cells, whole transcriptome sequencing, non-coding RNA, ceRNA network

OPEN ACCESS

Edited by:
Hua Li,
Shanghai Jiao Tong University, China

Reviewed by:
Lei Jiang,
Chinese Academy of Sciences, China
Chunxi Yang,
Shanghai JiaoTong University, China
Chuandong Wang,
Shanghai Jiaotong University, China

****Correspondence:***
Daping Wang
wangdp@mail.sustech.edu.cn

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *14 September 2021*
Accepted: *18 November 2021*
Published: *20 December 2021*

Citation:
Geng Y, Chen J, Chang C, Zhang Y, Duan L, Zhu W, Mou L, Xiong J and Wang D (2021) Systematic Analysis of mRNAs and ncRNAs in BMSCs of Senile Osteoporosis Patients. Front. Genet. 12:776984. doi: 10.3389/fgene.2021.776984

INTRODUCTION

Senile osteoporosis (SOP) is an age-related skeleton disease characterized by decreased bone mass and strength, which may lead to increased risk of fragility fractures (Raisz, 2005; Brown, 2017). With the increase of longevity, senile osteoporosis has recently become a major chronic metabolic bone disease in the world. Among people over 50 years old, one third of women and one fifth of men are susceptible to suffer from osteoporosis-induced bone fracture (Hendrickx et al., 2015). Bone homeostasis is a dynamic balance mediated by osteoblastic bone formation and osteoclastic bone resorption, and insufficient bone formation could lead to osteoporosis (Kim et al., 2020). Osteoblasts, which are differentiated from bone mesenchymal stem cells (BMSCs), play fundamental roles in bone formation by increasing the amount of matrix and the level of mineralization (Kiernan

et al., 2017). In senile, the proliferation and differentiation level of BMSCs significantly reduced, leading to bone formation deficit and osteoporosis (Kiernan et al., 2016). Therefore, it is of great important to investigate the molecular mechanism of impaired BMSCs in aging-induced osteoporosis and search for novel therapeutic targets.

Non-coding RNAs (ncRNAs) are non-protein coding transcripts served as crucial regulators of gene expression and cell fate (Schwarzer et al., 2017), which has been suggested as a novel class of potential diagnostic biomarkers and therapeutic targets (Beermann et al., 2016). The tremendous progress of high-throughput sequencing technologies has resulted in a plethora of studies focusing on ncRNAs, including microRNAs (miRNAs), circular RNAs (circRNAs), as well as long ncRNAs (lncRNAs) (Anastasiadou et al., 2018). As of today, the role of miRNAs in osteogenic differentiation of BMSCs have been widely explored, partially elucidating the pathogenesis of osteoporosis (Ell and Kang, 2014). LncRNAs, defined as linear transcripts of over 200 nucleotides, have also been suggested as therapeutic targets in various human diseases including osteoporosis (Kopp and Mendell, 2018; Chen et al., 2020; Del et al., 2020). Unlike abovementioned linear RNAs, circRNAs are characterized by a stable and closed RNA loops lacking 5′ and 3′ ends. Importantly, the expression of circRNAs is tissue- and cell-specific, indicating that circRNAs could involve in special biological pathways and thus exhibit unique cellular functions (Salzman et al., 2013). Indeed, both lncRNAs and circRNAs are able to act as competing endogenous RNAs (ceRNAs) by targeting miRNAs, and hence regulating the expression of miRNAs' target genes (Hansen et al., 2013; Tao et al., 2019). Recently, multiple studies have suggested that lncRNAs and circRNAs play significant roles in osteoporosis. For instance, lncRNA GAS5 has been revealed to promote osteogenic differentiation of mouse BMSCs by targeting the miR-135a-5p/FOXO1 axis (Wang et al., 2019). LncRNA Xist functions as a molecular sponge of miR-19a-3p to repress osteogenesis in age-related osteoporosis mouse model (Chen et al., 2020). Downregulation of hsa-circRNA 0006393 has been found in BMSCs from glucocorticoid-induced osteoporosis (GIOP) patients, which was shown to enhance osteogenesis by targeting miR-145-5p/FOXO1 (Wang et al., 2019). CircRNA 0016624 could regulate osteogenic differentiation of BMSCs by targeting miR-98 *in vitro* (Yu and Liu, 2019). However, transcriptome-wide profiling of abnormally expressed ncRNAs in senile osteoporosis as well as their regulatory networks have not been extensively investigated.

To screen the potential ncRNAs involved in senile osteoporosis and explore their molecular mechanism, we systematically analyze the expression profiles of miRNAs, lncRNAs, circRNAs and mRNAs in BMSCs of senile osteoporosis patients and normal individuals by whole transcriptome sequencing. According to the differentially expressed ncRNAs in the SOP group, we found that circRNA008876-miR-150-5p-mRNA represented the only distinctive ceRNA network, which was further validated by molecular biology experiments including qRT-PCR, western blot, overexpression and knockdown. In conclusion, our study established that circRNA008876-miR-150-5p-mRNA could be an important ceRNA network involved in senile osteoporosis, which provides potential biomarkers and therapeutic targets for senile osteoporosis.

MATERIALS AND METHODS

Isolation and Culture of Human BMSCs

Human BMSCs (hBMSCs) of senile osteoporosis were isolated from the bone marrow of discarded bone tissue from male senile osteoporosis patients (79 ± 5.6 years) who underwent joint replacement and diagnosed as SOP with bone minerality tests using Dual Energy X-ray Absorptiometry (Lorente et al., 2012). The normal hBMSCs were obtained using ScienCell (#7500), derived from male normal individuals under 20 years. The hBMSCs were cultured in MesenGro human mesenchymal stem cells medium (StemRD) supplemented with 10% FBS (E510008, Sangon Biotech) at 37°C with humidified 5% CO_2. The third passage of hBMSCs were utilized for whole transcriptome sequencing. DMSO was ordered from Sangon Biotech (A503039). The entire research plan has been approved by the ethics committee of Shenzhen Second People's Hospital (The First Hospital Affiliated to Shenzhen University).

Whole Transcriptome Sequencing

Three SOP patients-derived hBMSCs (BP1, BP2, and BP3) and three normal individuals (BN1, BN2, and BN3) derived hBMSCs were subjected to whole transcriptome sequencing. Total RNA was extracted from hBMSCs using Trizol reagent (Invitrogen) according to the manufacturer's instructions. After quantification and integrity of RNA samples were verified, rRNAs were removed from total RNA using Ribo-Zero rRNA Removal Kits (Illumina). Then, RNA libraries were constructed with TruSeq Stranded Total RNA Library Prep Kit (Illumina), and the two libraries were then sequenced on HiSeq-4000 (Illumina).

Construction of ceRNA Network

The miRNA binding regions of ncRNA and mRNA were analyzed using miRanda software (version 0.10.80) and Targetscan software (Release 7.2). The DE miRNAs targeting DE lncRNA, circRNA and mRNA were predicted to form the lncRNA/circRNA-miRNA-mRNA networks.

GO Enrichment and KEGG Pathway Analysis

Gene function related to senile osteoporosis was analyzed by GO annotation derived from Gene Ontology (www.geneontology.org) in which mRNA were grouped according to biological processes (BPs), cellular components (CCs) and molecular functions (MFs). The related pathways were enriched by KEGG and presented in a scatter plot diagram. We were only interested in biological processes and KEGG pathways showing significance according to the following parameters: $p < 0.05$, FDR <0.05, and enrichment score >1.5.

Real-Time Quantitative PCR

cDNA was synthesized with PrimeScript™ RT reagent Kit (Takara) using 1ug total RNA after gDNA was erased according to the

manufacturer's instructions. Applied Biosystems 7500 Real-Time PCR Systems was applied to perform amplification reaction using SYBR Premix Ex Taq (Takara). *GAPDH* and *U6* were used as endogenous controls for circRNA008876, osteogenic genes and miR-150-5p expressions, respectively. The RNA expressions were analyzed by $2^{-\triangle\triangle CT}$ method. The primer sequences were listed in **Supplementary Table S1**.

Dual-Luciferase Reporter Assay

HEK293T cells were co-transfected with recombinant plasmids (circRNA BR or circRNA BR Mut) and miR-337-3p mimics or negative control siRNAs (GenePharma). The recombinant plasmids were constructed with psiCHECK™-2 using the primers presented in **Supplementary Table S2**. Lipofectamine 2000 (Invitrogen) was used for transient transfection according to the manufacturer's instructions. After 24 h, cells were lysed and Firefly and Renilla luciferase activities were measured by Dual-Luciferase Reporter Assay System (Promega). Biological triplicates were performed.

ALP Staining

ALP staining was applied to test the osteogenesis of BMSCs after transfection of overexpressing vector or siRNA of circRNA008876 for 7 days according to the ALP staining kit protocol (Beyotime, C3206). Photographs were obtained under microscope (Nikon TE 2000).

Statistical Analysis

Statistical analyses were performed by the GraphPad Prism 8.0 and SPSS 22.0. Statistical significance of two or more biological replicates were performed by analysis of variance (ANOVA) and *post hoc* test. All results are presented as mean ± SD. $p < 0.05$ was considered statistically significant.

RESULT

Evaluation of hBMSCs Isolated From SOP Model

The hBMSCs isolated from SOP patients (passage 0) significantly presented a fibroblastic morphology similar to normal hBMSCs (**Figure 1A**). The surface markers of mesenchymal stem cells such as CD29, CD73 and CD105 were positively expressed (≥98%) in both groups of hBMSCs, while the hematopoietic markers CD34, CD45 and HLA-DR were negatively expressed (≤0.1%) in both BMSCs as evidenced by flow cytometry results (**Figure 1B**). Additionally, qPCR results showed that the expression levels of osteogenic genes including *RUNX2*, *ALPL*, *COL1A1* and *SPP1* were significantly reduced in BMSCs from SOP patients as compared to those from normal individuals, indicating the decreased osteogenic differentiation potential of BMSCs in SOP patients (**Figure 1C**).

Differentially Expressed lncRNAs, circRNA, miRNAs and mRNAs

We then compared the expression profiles of ncRNAs and mRNAs of hBMSCs from the two groups. With a fold change cutoff value >2 and p value <0.05, 30 lncRNAs (23 upregulated and 7 downregulated), 6 circRNAs (3 upregulated and 3 downregulated), 27 miRNAs (23 upregulated and 4 downregulated) and 415 mRNAs (60 upregulated and 51 downregulated) were identified as differentially expressed (DE) genes between hBMSCs from SOP patients and normal ones. The hierarchical cluster analysis of lncRNA, circRNA, miRNA, and mRNA demonstrated significant differences between hBMSCs from normal (BN) and SOP (BP) groups (**Figures 2A–D**). Complete information of all DE ncRNAs and mRNAs are provided in **Supplementary Information**. RNA-seq data were uploaded in Short Read Archive (SRA) of National Center for Biotechnology Information (NCBI) with accession number SRP337202 under Bioproject PRJNA763497.

Gene Ontology (GO) Enrichment and KEGG Pathway Analysis

We then performed GO enrichment analysis to reveal the key regulators and vital pathways potentially engaged by DE ncRNAs and mRNA in SOP. The top enriched terms regarding biological process (BP), cellular component (CC), and molecular function (MF) of four categories of RNA were shown in **Figures 3A–D**, respectively. A quick survey on the bioinformatics analysis results revealed that several terms were highly related to osteogenesis. For example, the terms "response to growth factor," "cell adhesion" and "regulation of migration" are involved in the proliferation of BMSCs, which were remarkably enriched as top biological processes engaged by DE ncRNAs (**Figure 3A**); similar terms, including "regulation of growth" and "homeostatic process," were engaged by circRNA-located genes (**Figure 3B**). For DE miRNA-targeted genes, key processes related to cell division were enriched, including the terms "gene expression," "DNA integrity checkpoint" and "mitotic DNA damage checkpoint" (**Figure 3C**)". For DE mRNAs, three BP terms, namely "regulation of cell migration," "CXCR chemokine receptor binding," and "insulin-like growth factor binding," were related to proliferation and differentiation of BMSCs (**Figure 3D**).

We also utilized the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to categorize the DE ncRNAs-targeted genes and mRNAs. The top 20 pathways of KEGG analysis for each category of DE ncRNAs or mRNAs were shown in **Figures 4A–D**. Among the functional pathways associated with DE miRNAs-targeted genes, we noticed that "TGF-beta signaling pathway" and "cell adhesion molecules (CAMs)" were highly related to osteogenesis (**Figure 4C**), while "CAMs" were also significantly engaged by DE mRNAs (**Figure 4D**). In addition, the terms "cytokine-cytokine receptor interaction", "chemokine signaling pathway" and "mineral absorption" associated with DE mRNAs were also related to osteogenic differentiation of BMSCs (**Figure 4D**).

Construction and Analyses of ceRNA Networks

Both lncRNA and circRNA are known to sponge miRNAs to prevent their interactions with target mRNAs, thus exhibiting

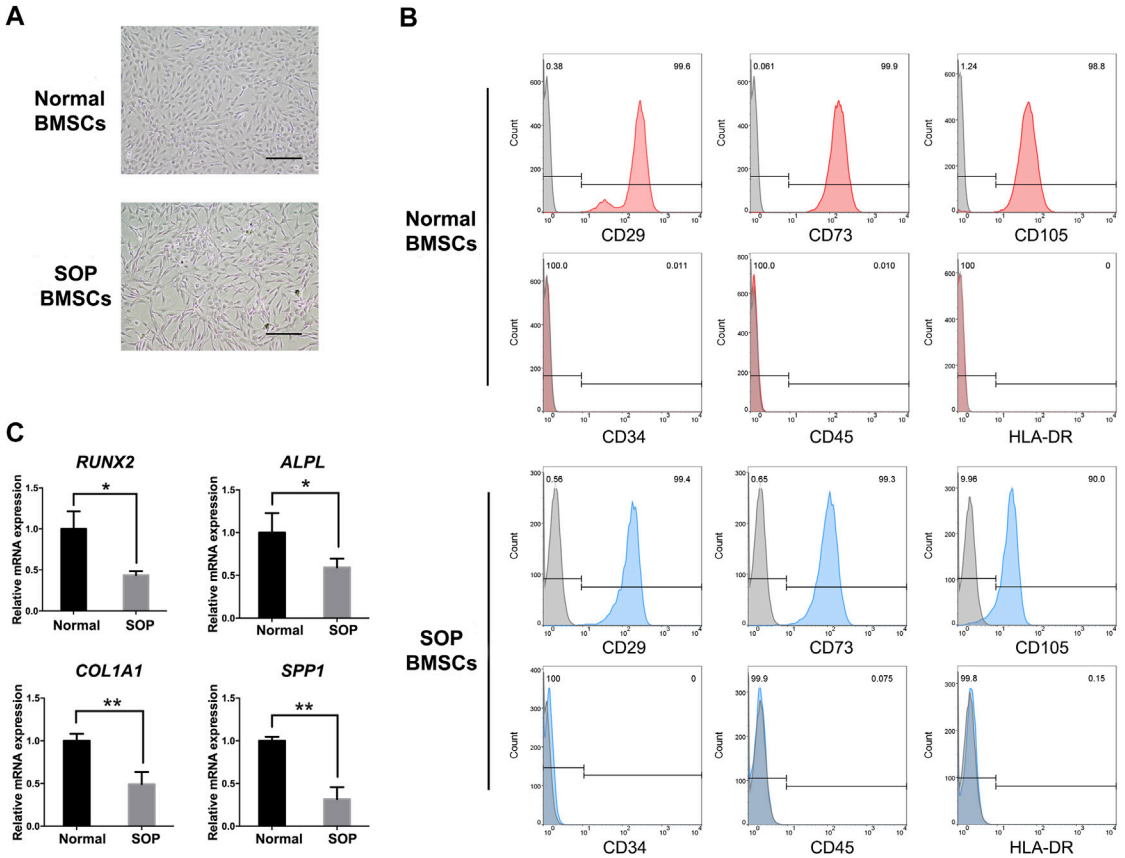


FIGURE 1 | Characterization of hBMSCs from senile osteoporosis patient. **(A)** The morphology of P2 hBMSCs from normal individuals (Normal BMSCs) and SOP patients (SOP BMSCs); **(B)** Stem cell surface markers of hBMSCs were analyzed by flowcytometry. Red (Normal BMSCs) and blue curves (SOP BMSCs) presented positive (CD29, CD73, and CD105) and negative markers (CD34, CD45 and HLA-DR) of hBMSCs, respectively, while grey curves represent isotype controls; **(C)** Expression of osteogenic genes *RUNX2, ALPL, COL1a1* and *SPP1* of hBMSCs from SOP patients and control group were measured by qPCR. *GAPDH* expression was used as an internal control for mRNA expression. Scale bar: 200 μm. Error bars, SEM (n = 3). $^{*}p < 0.05$, $^{**}p < 0.01$.

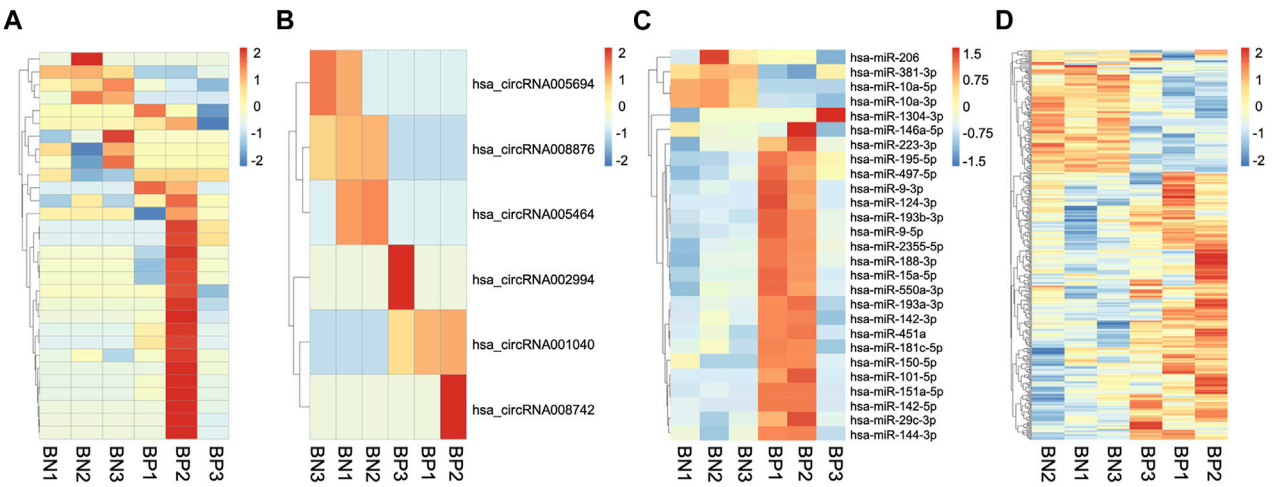


FIGURE 2 | Expression profiles of mRNAs and ncRNAs. Heatmaps clustering of all significantly DE lncRNAs **(A)**, circRNAs **(B)**, miRNAs **(C)** and mRNAs **(D)** in hBMSCs of three senile osteoporosis patients (BP1, BP2, and BP3) and three normal individuals (BN1, BN2, and BN3).

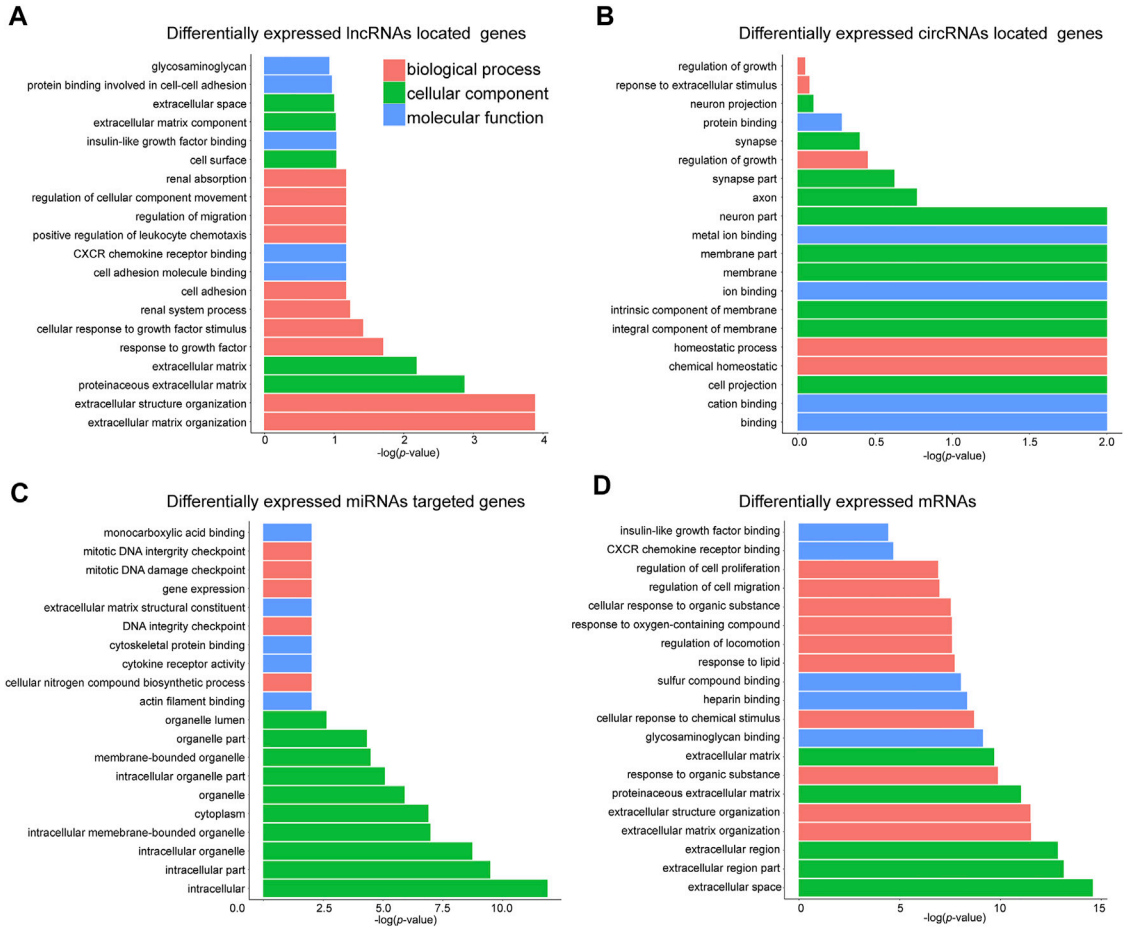


FIGURE 3 | GO analysis of DE mRNAs, lncRNAs, circRNAs and miRNAs. Top 20 terms of Gene Ontology (GO) enrichment analysis regarding **(A)** DE lncRNAs-located genes; **(B)** DE circRNAs-located genes; **(C)** DE miRNAs-targeted genes; and **(D)** DE mRNAs.

competitive endogenous RNA (ceRNA) activity. To predict the functions of DE ncRNAs and mRNAs and uncover the potential ceRNA networks regulating the osteogenic differentiation of hBMSCs in SOP patients, lncRNA-miRNA-mRNA and circRNA-miRNA-mRNA networks were constructed by using Targetscan and Miranda softwares.

As shown in **Figure 5A**, six lncRNAs were predicted as miRNA sponges, which forms five networks including lncRNA_00214189/miR-206, lncRNA_00056143/miR-381-3p, lncRNA_00211178/miR-2355-5p, lncRNA_00054644/miR-181c-5p, lncRNA_00220556/miR-181c-5p and lncRNA_00206603/miR-150-5p. Among the miRNA-targeted genes in these networks, a number of osteogenesis-related genes were identified including several upregulated genes CCL2, WISP1, WISP2, and one downregulated gene FBN2 (Si et al., 2006; Ono et al., 2011; Smaldone et al., 2011; Córdova et al., 2017). To further understand the biological function of these lncRNA-miRNA-mRNA networks, GO enrichment and KEGG pathway analyses were performed on the target genes. The top 5 cellular component terms were all related to extracellular matrix (**Figure 5B**), which were highly correlated with the GO enrichment analyses results of DE lncRNAs-located genes and mRNAs (**Figure 3**). The top biological process terms, such as "response to hypoxia," "cell migration," and "cell motility," are obviously involved in osteogenic differentiation of BMSCs. Moreover, the term "cytokine-cytokine receptor interaction" represents the most remarkable KEGG pathway in these lncRNA-constructed ceRNA networks (**Figure 5C**), which was also significantly enriched in the KEGG analysis results for DE mRNAs.

In the case of circRNA, only one ceRNA network, namely circRNA008876/miR-150-5p, was constructed (**Figure 6A**). Among the target genes of miR-150-5p, HOXB3 and HHIP have been reported to associate with osteogenesis. GO analysis of target genes of this network showed that "mRNA translation and transport" was the most significant biological function in hBMSCs (**Figure 6B**). In addition, "p53 signaling pathway" and "Hedgehog signaling pathway" were significantly enriched, which have also been suggested to play important roles in osteogenesis (Tyner et al., 2002; Onodera et al., 2020) (**Figure 6C**). These results further suggested that both

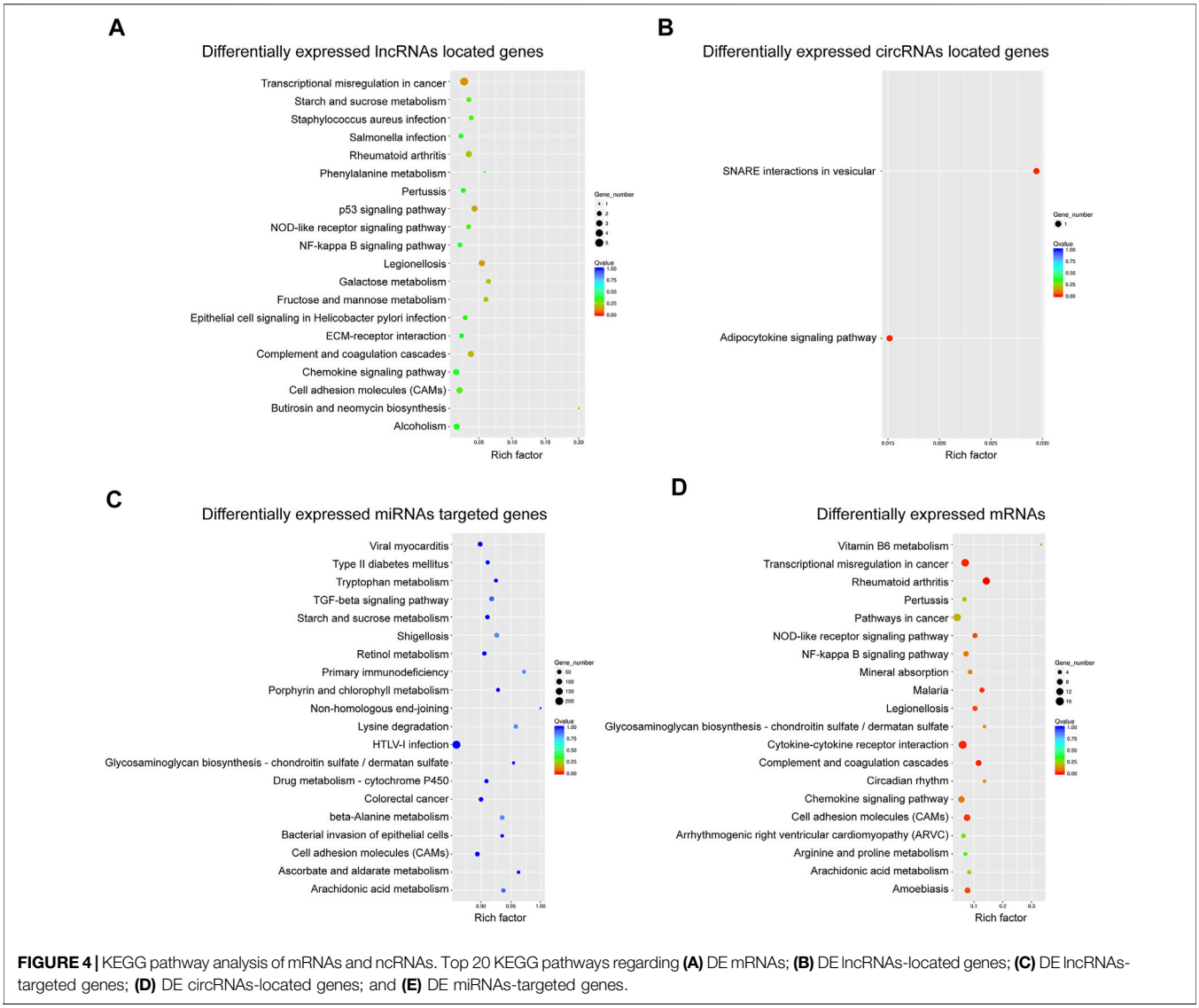


FIGURE 4 | KEGG pathway analysis of mRNAs and ncRNAs. Top 20 KEGG pathways regarding **(A)** DE mRNAs; **(B)** DE lncRNAs-located genes; **(C)** DE lncRNAs-targeted genes; **(D)** DE circRNAs-located genes; and **(E)** DE miRNAs-targeted genes.

lncRNAs and circRNAs could act as miRNA sponge to regulate the expression of downstream genes, which in turn take part in osteogenic differentiation of BMSCs. Importantly, circRNA008876/miR-150-5p represents a unique circular RNA-based ceRNA network in senile osteoporosis. Therefore, we further explored the biological function and validated the role of circRNA008876 in osteogenic differentiation of hBMSCs.

Validation of circRNA008876/miR-150-5p ceRNA Network in SOP

To validate the connection between circRNA00876 and the clinicopathological characteristics of senile osteoporosis, we performed RT-qPCR experiments and the results suggested that reduced expression of circRNA008876 was significantly detected in SOP (**Figure 7A**), and accompanied with the elevated expression level of miR-150-5p (**Figure 7B**). To validate the association between circRNA00876 and miR-150-5p, at first, we predicted the possible binding sites (**Figure 7C**). Then, a dual-luciferase reporter assay was performed by constructing luciferase reporter plasmids harboring the predicted binding sequence of circRNA008876 and its mutation. The luciferase activity of the vector carrying the wild-type binding site of circRNA008876 was significantly inhibited by miR-150-5p mimics, while the luciferase activity did not change in binding site mutated group (**Figure 7D**), suggesting the direct interaction between circRNA008876 and miR-150-5p. To further confirm the role of circRNA008876/miR-150-5p network in osteogenic differentiation of hBMSCs, we investigated the individual effect of circRNA008876 or miR-150-5p on the expression of several osteogenesis-related genes, including *RUNX2, ALPL, COL1A1, SPP1*, and *OSX* by qPCR. Overexpression of circRNA008876 significantly upregulated the

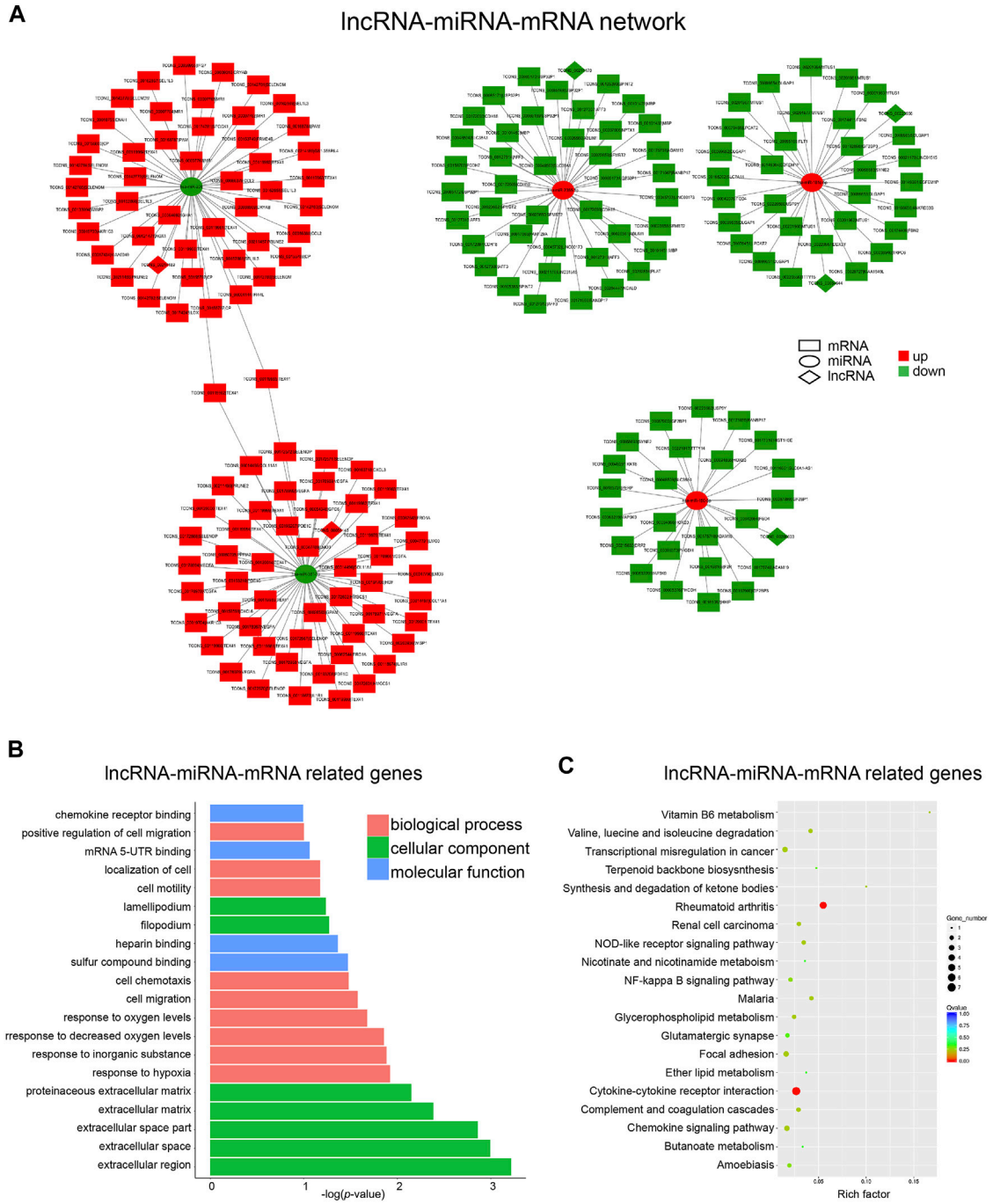


FIGURE 5 | Construction of the lncRNA-miRNA-mRNA Network. **(A)** Network analysis of lncRNA (diamond)-miRNA (round)-mRNA (square). Red and green represent up- and down-regulation, respectively. **(B)** Top 20 GO terms analyzed with mRNAs in lncRNA-miRNA-mRNA network. **(C)** Top 20 KEGG pathway enrichment of mRNAs in lncRNA-miRNA-mRNA network.

mRNA level of the osteogenic genes, while inhibition of circRNA008876 decreased their mRNA level (**Figure 7E**). The effect of circRNA008876 on osteogenic differentiation was further confirmed by ALP staining (**Figure 7G**). Accordingly, up- or down-regulation of miR-150-5p exhibited opposite effect as that of circRNA008876 (**Figure 7F**), further indicating miR-150-5p as the downstream gene of circRNA008876 in SOP. Taken together, these results demonstrated that circRNA008876 promotes osteogenesis via sponging miR-150-5p, and reduced expression of circRNA008876 leads to impaired osteogenesis in senile osteoporosis through miR-150-5p.

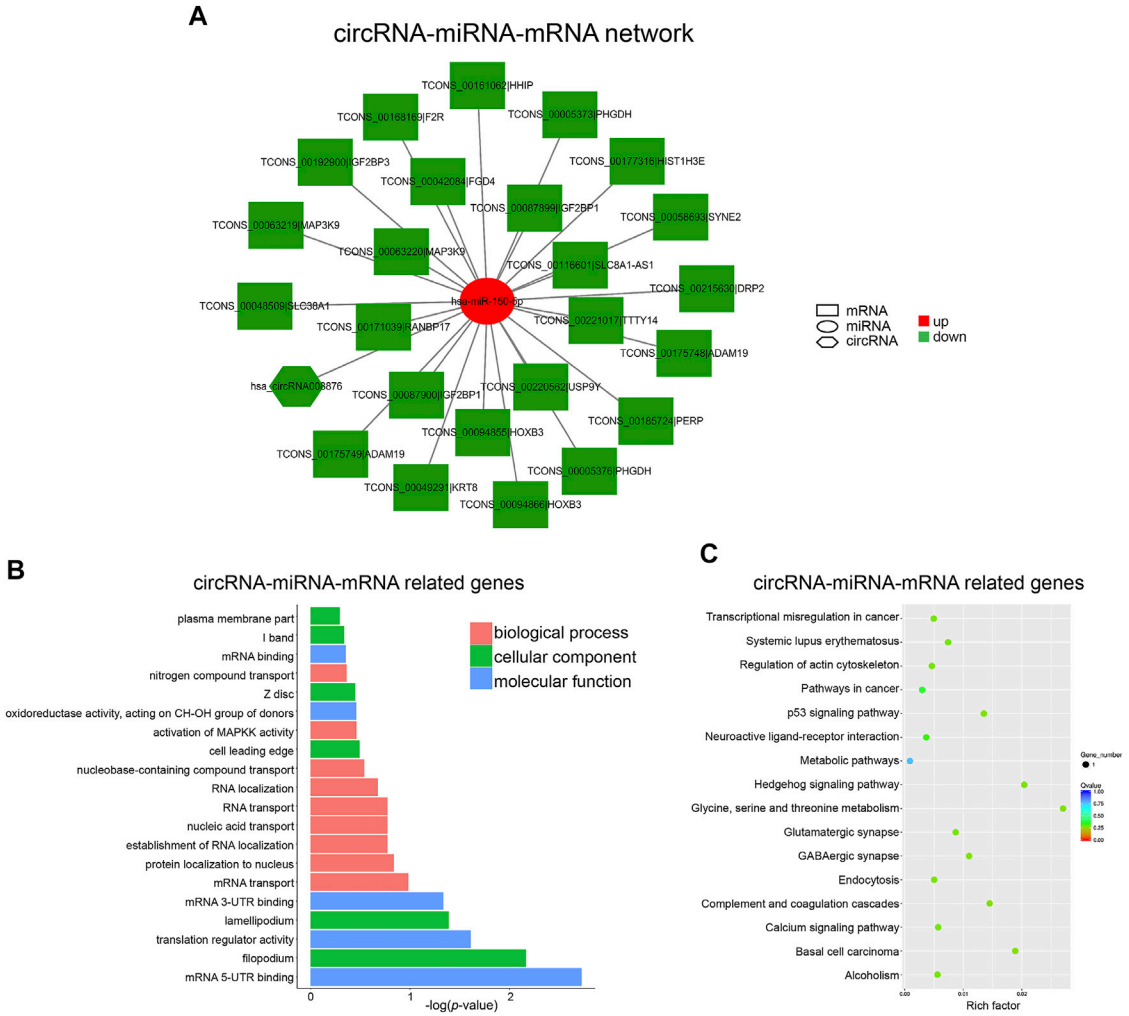


FIGURE 6 | Construction of the circRNA-miRNA-mRNA Network. **(A)** circRNA-miRNA-mRNA network was constructed with down-regulated circRNA_008876 and up-regulated miR-150-5p. **(B)** Top 20 GO terms analyzed with mRNAs in circRNA-miRNA-mRNA network. **(C)** Top 20 KEGG pathway enrichment of mRNAs in circRNA-miRNA-mRNA network.

DISCUSSION

Osteoporosis is a complex bone disease caused by multiple pathogenetic mechanisms, leading to skeletal microarchitectural deterioration and bone mass loss. The pathophysiology and molecular mechanism of senile (type II) osteoporosis (SOP) is different from estrogen deficiency-induced postmenopausal (type I) osteoporosis (PMOP) caused by the activation of osteoclasts. Whereas the osteogenesis deficiency of hBMSCs represents an important pathogenic factor in senile osteoporosis (Rachner et al., 2011), the underlying mechanism remains elusive. Recent studies have shown that ncRNAs including miRNAs, lncRNAs and circRNAs, were significantly involved in osteoporosis development (Jin et al., 2018). Nevertheless, systematic analysis of DE ncRNAs and mRNAs, as well as the ceRNA regulatory networks in hBMSCs from SOP patients, have not be reported. In the present study, hBMSCs from senile osteoporosis patients were isolated, and its decreased osteogenic ability were validated by qPCR. By comparing hBMSCs from SOP patients and normal individuals, 30 lncRNAs, 6 circRNAs, 27 miRNAs, and 415 mRNAs were differentially expressed as suggested by the results of high-throughput RNA-seq.

The function of miRNA in osteogenesis has been extensively investigated in previous studies. Several up-regulated miRNAs of SOP in our study including miR-124-3p, miR-193a, miR-451a, and miR-144-3p have been proven as negative regulators in osteogenesis (Qadir et al., 2015; Lu et al., 2019; Tang et al., 2019; Li et al., 2020; Song et al., 2021). Further studies are needed to investigate the function of other DE miRNAs in SOP. On the other hand, the osteogenic regulatory function of lncRNAs, mainly by targeting and sponging miRNAs, has been reported in recent years.

For example, lncRNA MEG3-miR-133a-3p axis has been suggested to be involved in PMOP through inhibiting osteogenic differentiation of BMSCs (Wang et al., 2017). In addition, lncRNA

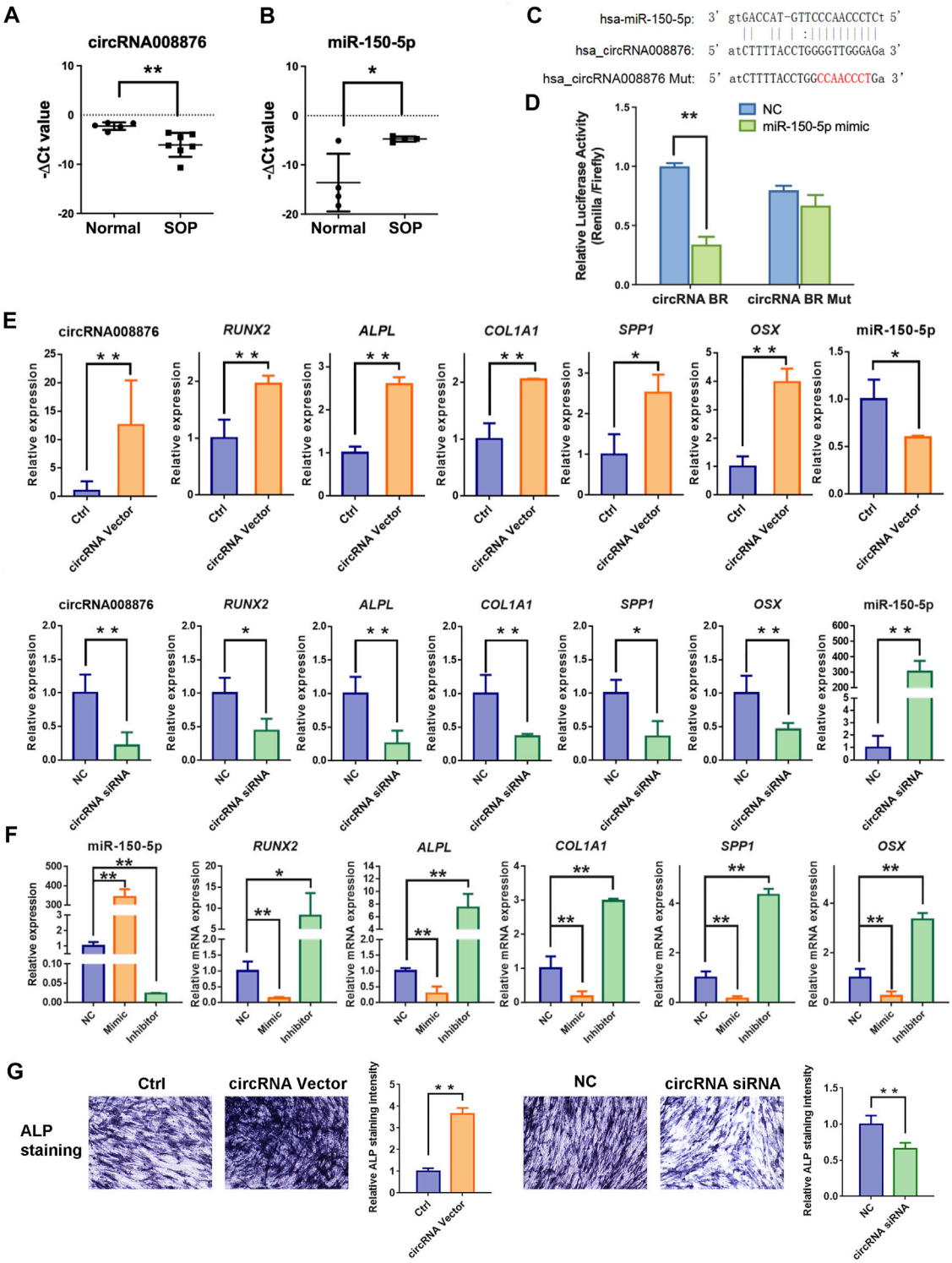


FIGURE 7 | circRNA008876/miR-150-5p regulate osteogenesis of hBMSCs. **(A–B)** Real-time PCR analysis of circRNA008876 **(A)** and miR-150-5p **(B)** expression of hBMSCs from SOP patients or the control group. *GAPDH* and *U6* expression were used as an internal control for circRNA008876 and miR-150-5p, respectively. **(C)** Schematic representation of the wild-type and mutant binding sites of circRNA008876 targeting miR-150-5p. **(D)** HEK293T cells were transfected with psiCHECK™-2 Vector containing a fragment of circRNA008876 binding sites for miR-150-5p, or the corresponding mutant constructs. The effect of miR-150-5p mimics on the corresponding vector luciferase activity was tested. **(E)** hBMSCs were transfected with circRNA008876 overexpressing vector or silencing siRNA. Expression of circRNA008876, miR-150-5p and osteogenesis genes including *RUNX2, ALPL, COL1a1*, *SPP1* and *OSX* were tested by qPCR 3 days after transfection. (F) hBMSCs were transfected with miR-150-5p mimic or inhibitor. Expression of miR-150-5p and osteogenesis genes were tested by qPCR 3 days after transfection. (G) Alkaline phosphatase (ALP) staining on Day 7 of hBMSCs transfected with circRNA008876 overexpressing vector or silencing siRNA. Images were captured under microscope with 100X enlargement. Relative ALP staining intensity were analyzed by ImageJ. Error bars, SEM (n = 3). $^{*}p < 0.05$, $^{**}p < 0.01$.

H19 was found to be a positive regulator of osteogenesis in PMOP via miR-532-3p/SIRT1 axis (Li et al., 2021). In age-related osteoporosis, lncRNA Xist has been reported to inhibit osteogenesis by targeting and sponging miR-19a-3p (Chen et al., 2020). Nevertheless, global investigation and functional studies of lncRNAs in senile osteoporosis is still lacking. In this study, we surprisingly found that the expression of lncRNA_00218,705|XIST also increased in SOP patients' BMSCs. On the other hand, we revealed several additional significantly up-regulated lncRNAs including lncRNA_00218,705|XIST, lncRNA_00090434|AC144831.1, and lncRNA_00214,189|GS1-358P8.4, as well as several down-regulated lncRNAs such as lncRNA_00211,178|LINC01410 and lncRNA_00219,655|MIR503HG in SOP patients. These lncRNAs could be potential clinical biomarkers which need further validation. Recently, the role of circRNAs in osteoporosis and osteogenesis have also been investigated. For example, has_circ_0001275 has been proven to be a potential diagnostic biomarker for PMOP (Zhao et al., 2018), and some DE circRNAs have been detected in various biological systems, including rBMSCs (Li et al., 2017), human periodontal ligament stem cells (Zheng et al., 2017), and MC3T3-E1 cells (Qian et al., 2017). Unfortunately, the underlying mechanism has yet not been fully explored. Moreover, global profiling of DE circRNAs in senile osteoporosis has not been reported yet. Here, we successfully identified 6 novel DE circRNAs that may take part in senile osteoporosis, all of which have not been studied in osteoporosis before.

The functions of DE ncRNAs and mRNAs in senile osteoporosis pathology was pedicted by GO enrichment (BP, CC, and MF) and KEGG pathways analyses. As we known, extracellular matrix mineralization is an essential progress in bone formation (Murshed, 2018). According to the results of GO enrichment analysis, we noticed that extracellular matrix is one the most significant terms of CC analysis regarding DE mRNAs, DE lncRNAs-located mRNAs, and DE miRNAs-targeted mRNAs. Meanwhile, CXCR chemokine receptor binding were significantly enriched in MF analysis regarding both DE mRNAs and DE lncRNAs-located mRNAs, validating previous results regarding the role of CXCR-4 in recruiting MSCs in bone repair (Zhang et al., 2017). Moreover, KEGG pathway analysis revealed that the adipocytokine signaling pathway was significantly engaged by DE circRNAs-located mRNAs, which implies that DE circRNAs may participate in the balance of osteogenesis and adipogenesis of hBMSCs. It is well-established that the shift from osteogenesis to adipogenesis is a major pathological progress of SOP (Chen et al., 2016). Noticeably, lncRNAs may regulate p53 signaling to affect hBMSCs differentiation, as p53 has been proved to promote osteogenesis in mesenchymal stem cells (Tataria et al., 2006).

In general, the regulatory function of cell differentiation by ncRNAs are mainly through the ceRNA network. In this study, five lncRNA-miRNA-mRNA networks were constructed, among which miR-206, miR-381-3p were down-regulated, while miR-2355-5p, miR-181c-5p, and miR-150-5p were up-regulated. Previous reports suggested that miR-206 and miR-381-3p were able to inhibit osteogenic differentiation of BMSCs (Chen et al., 2019; Long et al., 2019), which were in conflict with our results. The effect of miR-2335-5p, miR-101c-5p in osteogenic differentiation has never been reported. In addition, miR-150, sponged by lncRNA TCONS_00206,603, was proved to serve as a negative regulator for osteoblasts (Moussa et al., 2021), which was found to be upregulated in SOP patients in our study. However, we have not detected any downregulation of lncRNA TCONS_00206,603 in SOP patient. Interestingly, down-regulation of circRNA008876 was predicted to be a unique circRNA sponge for miR-150-5p. Thus, we systematically validated the role of circRNA008876 in BMSCs osteogenesis by targeting and sponging miR-150-5p. However, there are some limitations in this study. That is, the molecules downstream of the circRNA008876/miR-150-5p axis are unknown, and *in vivo* pro-osteogenesis effect induced by circRNA008876 overexpression is unclear. To be noted, the abovementioned genes identified in this study, mainly including lncRNAs and circRNAs, did not overlap in previous studies. These results can be attributed to the differences in patients and ethnicity, which need further investigation.

CONCLUSION

In the present study, we systematically explored the role of differently expressed ncRNAs and mRNAs in hBMSCs isolated from senile osteoporosis patients by whole transcriptome sequencing. The ceRNA regulatory networks of lncRNA/circRNA-miRNA-mRNA were constructed to further investigate the function of ncRNAs in senile osteoporosis. Among them, circRNA008876 was identified as a unique and critical circRNA in the occurrence and development of senile osteoporosis and might serve as a potential diagnostic biomarker for senile osteoporosis. Overall, our findings provide an insight into the molecular mechanisms of senile osteoporosis in the view of ncRNAs, which may offer novel therapeutic targets for clinic treatment of senile osteoporosis.

DATA AVAILABILITY STATEMENT

The RNA sequencing datasets can be found in SRA of NCBI with accession number SRP337202 under BioProject PRJNA763497. The data are accessible with the following link: https://www.ncbi.nlm.nih.gov/sra/PRJNA763497.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of The Shenzhen Second People's Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YG and DW designed the experiments; JC, CC, and YZ conducted part of the cell experiments; LM and WZ helped to collect BMSCs samples; YG, LD, and JX analyzed the results and wrote the manuscript; DW reviewed the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (No. 81972116; No. 81772394); Natural Science Foundation of Guangdong Province (No. 2019A1515011108); Guangdong International Cooperation Project (No. 2021A0505030011); Key Program of Natural Science Foundation of Guangdong Province (No. 2018B0303110003); and Shenzhen Science and Technology Projects (No. GJHZ20200731095606019; No. JCYJ20170817172023838; No. JCYJ20170413161649437).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.776984/full#supplementary-material

REFERENCES

Anastasiadou, E., Jacob, L. S., and Slack, F. J. (2018). Non-coding RNA Networks in Cancer. *Nat. Rev. Cancer* 18, 5–18. doi:10.1038/nrc.2017.99

Beermann, J., Piccoli, M.-T., Viereck, J., and Thum, T. (2016). Non-coding RNAs in Development and Disease: Background, Mechanisms, and Therapeutic Approaches. *Physiol. Rev.* 96, 1297–1325. doi:10.1152/physrev.00041.2015

Brown, C. (2017). Staying strong. *Nature* 550, S15–S17. doi:10.1038/550S15a

Chen, J., Wang, Y., Wang, C., Hu, J.-F., and Li, W. (2020). LncRNA Functions as a New Emerging Epigenetic Factor in Determining the Fate of Stem Cells. *Front. Genet.* 11, 277. doi:10.3389/fgene.2020.00277

Chen, Q., Shou, P., Zheng, C., Jiang, M., Cao, G., Yang, Q., et al. (2016). Fate Decision of Mesenchymal Stem Cells: Adipocytes or Osteoblasts? *Cell Death Differ.* 23, 1128–1139. doi:10.1038/cdd.2015.168

Chen, S., Li, Y., Zhi, S., Ding, Z., Huang, Y., Wang, W., et al. (2020). lncRNA Xist Regulates Osteoblast Differentiation by Sponging miR-19a-3p in Aging-Induced Osteoporosis. *Aging Dis.* 11, 1058–1068. doi:10.14336/AD.2019.0724

Chen, Y., Yang, Y., Fan, X., Lin, P., Yang, H., Chen, X., et al. (2019). miR-206 Inhibits Osteogenic Differentiation of Bone Marrow Mesenchymal Stem Cells by Targetting Glutaminase. *Biosci. Rep.* 39, BSR20181108. doi:10.1042/BSR20181108

Córdova, L. A., Loi, F., Lin, T.-H., Gibon, E., Pajarinen, J., Nabeshima, A., et al. (2017). CCL2, CCL5, and IGF-1 Participate in the Immunomodulation of Osteogenesis during M1/M2 Transition *In Vitro*. *J. Biomed. Mater. Res.* 105, 3069–3076. doi:10.1002/jbm.a.36166

Del Real, A., López-Delgado, L., Sañudo, C., García-Ibarbia, C., Laguna, E., Perez-Campo, F. M., et al. (2020). Long Noncoding RNAs as Bone Marrow Stem Cell Regulators in Osteoporosis. *DNA Cel Biol.* 39, 1691–1699. doi:10.1089/dna.2020.5672

Ell, B., and Kang, Y. (2014). MicroRNAs as Regulators of Bone Homeostasis and Bone Metastasis. *Bonekey Rep.* 3, 549. doi:10.1038/bonekey.2014.44

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA Circles Function as Efficient microRNA Sponges. *Nature* 495, 384–388. doi:10.1038/nature11993

Hendrickx, G., Boudin, E., and Van Hul, W. (2015). A Look behind the Scenes: the Risk and Pathogenesis of Primary Osteoporosis. *Nat. Rev. Rheumatol.* 11, 462–474. doi:10.1038/nrrheum.2015.48

Jin, D., Wu, X., Yu, H., Jiang, L., Zhou, P., Yao, X., et al. (2018). Systematic Analysis of lncRNAs, mRNAs, circRNAs and miRNAs in Patients with Postmenopausal Osteoporosis. *Am. J. Transl. Res.* 10 (5), 1498–1510.

Kiernan, J., Davies, J. E., and Stanford, W. L. (2017). Concise Review: Musculoskeletal Stem Cells to Treat Age-Related Osteoporosis. *Stem Cell Translational Med.* 6, 1930–1939. doi:10.1002/sctm.17-0054

Kiernan, J., Hu, S., Grynpas, M. D., Davies, J. E., and Stanford, W. L. (2016). Systemic Mesenchymal Stromal Cell Transplantation Prevents Functional Bone Loss in a Mouse Model of Age-Related Osteoporosis. *Stem Cell Transl. Med.* 5, 683–693. doi:10.5966/sctm.2015-0231

Kim, J.-M., Lin, C., Stavre, Z., Greenblatt, M. B., and Shim, J.-H. (2020). Osteoblast-Osteoclast Communication and Bone Homeostasis. *Cells* 9, 2073. doi:10.3390/cells9092073

Kopp, F., and Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell* 172, 393–407. doi:10.1016/j.cell.2018.01.011

Li, N., Liu, L., Liu, Y., Luo, S., Song, Y., and Fang, B. (2020). miR-144-3p Suppresses Osteogenic Differentiation of BMSCs from Patients with Aplastic Anemia through Repression of TET2. *Mol. Ther. - Nucleic Acids* 19, 619–626. doi:10.1016/j.omtn.2019.12.017

Li, T., Jiang, H., Li, Y., Zhao, X., and Ding, H. (2021). Estrogen Promotes lncRNA H19 Expression to Regulate Osteogenic Differentiation of BMSCs and Reduce Osteoporosis via miR-532-3p/SIRT1 axis. *Mol. Cell Endocrinol.* 527, 111171. doi:10.1016/j.mce.2021.111171

Li, X., Peng, B., Zhu, X., Wang, P., Xiong, Y., Liu, H., et al. (2017). Changes in Related Circular RNAs Following ERbeta Knockdown and the Relationship to rBMSC Osteogenesis. *Biochem. Biophys. Res. Commun.* 493, 100–107. doi:10.1016/j.bbrc.2017.09.068

Long, H., Zhu, Y., Lin, Z., Wan, J., Cheng, L., Zeng, M., et al. (2019). miR-381 Modulates Human Bone Mesenchymal Stromal Cells (BMSCs) Osteogenesis via Suppressing Wnt Signaling Pathway during Atrophic Nonunion Development. *Cell Death Dis.* 10, 470. doi:10.1038/s41419-019-1693-z

Lorente Ramos, R. M., Azpeitia Armán, J., Arévalo Galeano, N., Muñoz Hernández, A., García Gómez, J. M., and Gredilla Molinero, J. (2012). Dual Energy X-ray Absorptimetry: Fundamentals, Methodology, and Clinical Applications. *Radiología (English Edition)* 54, 410–423. doi:10.1016/j.rx.2011.09.023

Lu, X.-D., Han, W.-X., and Liu, Y.-X. (2019). Suppression of miR-451a Accelerates Osteogenic Differentiation and Inhibits Bone Loss via Bmp6 Signaling during Osteoporosis. *Biomed. Pharmacother.* 120, 109378. doi:10.1016/j.biopha.2019.109378

Moussa, F. M., Cook, B. P., Sondag, G. R., DeSanto, M., Obri, M. S., McDermott, S. E., et al. (2021). The Role of miR-150 Regulates Bone Cell Differentiation and Function. *Bone* 145, 115470. doi:10.1016/j.bone.2020.115470

Murshed, M. (2018). Mechanism of Bone Mineralization. *Cold Spring Harb. Perspect. Med.* 8, a031229. doi:10.1101/cshperspect.a031229

Ono, M., Inkson, C. A., Kilts, T. M., and Young, M. F. (2011). WISP-1/CCN4 Regulates Osteogenesis by Enhancing BMP-2 Activity. *J. Bone Miner. Res.* 26, 193–208. doi:10.1002/jbmr.205

Onodera, S., Saito, A., Hojo, H., Nakamura, T., Zujur, D., Watanabe, K., et al. (2020). Hedgehog Activation Regulates Human Osteoblastogenesis. *Stem Cel Rep.* 15, 125–139. doi:10.1016/j.stemcr.2020.05.008

Qadir, A. S., Um, S., Lee, H., Baek, K., Seo, B. M., Lee, G., et al. (2015). miR-124 Negatively Regulates Osteogenic Differentiation and *In Vivo* Bone Formation of Mesenchymal Stem Cells. *J. Cel. Biochem.* 116, 730–742. doi:10.1002/jcb.25026

Qian, D., Yan, G., Bai, B., Chen, Y., Zhang, S., Yao, Y., et al. (2017). Differential circRNA Expression Profiles during the BMP2-Induced Osteogenic Differentiation of MC3T3-E1 Cells. *Biomed. Pharmacother.* 90, 492–499. doi:10.1016/j.biopha.2017.03.051

Rachner, T. D., Khosla, S., and Hofbauer, L. C. (2011). Osteoporosis: Now and the Future. *The Lancet* 377, 1276–1287. doi:10.1016/S0140-6736(10)62349-5

Raisz, L. G. (2005). Pathogenesis of Osteoporosis: Concepts, Conflicts, and Prospects. *J. Clin. Invest.* 115, 3318–3325. doi:10.1172/JCI27071

Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L., and Brown, P. O. (2013). Cell-type Specific Features of Circular RNA Expression. *Plos Genet.* 9, e1003777. doi:10.1371/journal.pgen.1003777

Schwarzer, A., Emmrich, S., Schmidt, F., Beck, D., Ng, M., Reimer, C., et al. (2017). The Non-coding RNA Landscape of Human Hematopoiesis and Leukemia. *Nat. Commun.* 8, 218. doi:10.1038/s41467-017-00212-4

Si, W., Kang, Q., Luu, H. H., Park, J. K., Luo, Q., Song, W.-X., et al. (2006). CCN1/Cyr61 Is Regulated by the Canonical Wnt Signal and Plays an Important Role in Wnt3A-Induced Osteoblast Differentiation of Mesenchymal Stem Cells. *Mol. Cel. Biol.* 26, 2955–2964. doi:10.1128/MCB.26.8.2955-2964.2006

Smaldone, S., Carta, L., and Ramirez, F. (2011). Establishment of Fibrillin-Deficient Osteoprogenitor Cell Lines Identifies Molecular Abnormalities Associated with Extracellular Matrix Perturbation of Osteogenic Differentiation. *Cell. Tissue Res.* 344, 511–517. doi:10.1007/s00441-011-1167-9

Song, C.-Y., Guo, Y., Chen, F.-Y., and Liu, W.-G. (2021). Resveratrol Promotes Osteogenic Differentiation of Bone Marrow-Derived Mesenchymal Stem Cells through miR-193a/SIRT7 Axis. *Calcif. Tissue Int.* doi:10.1007/s00223-021-00892-7

Tang, J. Z., Lin, X., Zhong, J. Y., Xu, F., Wu, F., Liao, X. B., et al. (2019). miR-124 R-egulates the O-steogenic D-ifferentiation of B-one M-arrow-derived M-esenchymal S-tem C-ells by T-argeting Sp7. *Mol. Med. Rep.* 19, 3807–3814. doi:10.3892/mmr.2019.10054

Tao, L., Yang, L., Huang, X., Hua, F., and Yang, X. (2019). Reconstruction and Analysis of the lncRNA-miRNA-mRNA Network Based on Competitive Endogenous RNA Reveal Functional lncRNAs in Dilated Cardiomyopathy. *Front. Genet.* 10, 1149. doi:10.3389/fgene.2019.01149

Tataria, M., Quarto, N., Longaker, M. T., and Sylvester, K. G. (2006). Absence of the P53 Tumor Suppressor Gene Promotes Osteogenesis in Mesenchymal Stem Cells. *J. Pediatr. Surg.* 41, 624–632. doi:10.1016/j.jpedsurg.2005.12.001

Tyner, S. D., Venkatachalam, S., Choi, J., Jones, S., Ghebranious, N., Igelmann, H., et al. (2002). p53 Mutant Mice that Display Early Ageing-Associated Phenotypes. *Nature* 415, 45–53. doi:10.1038/415045a

Wang, Q., Li, Y., Zhang, Y., Ma, L., Lin, L., Meng, J., et al. (2017). LncRNA MEG3 Inhibited Osteogenic Differentiation of Bone Marrow Mesenchymal Stem Cells from Postmenopausal Osteoporosis by Targeting miR-133a-3p. *Biomed. Pharmacother.* 89, 1178–1186. doi:10.1016/j.biopha.2017.02.090

Wang, X. B., Li, P. B., Guo, S. F., Yang, Q. S., Chen, Z. X., Wang, D., et al. (2019). circRNA_0006393 Promotes Osteogenesis in Glucocorticoid-induced O-steoporosis by S-ponging miR-145-5p and U-pregulating FOXO1. *Mol. Med. Rep.* 20, 2851–2858. doi:10.3892/mmr.2019.10497

Wang, X., Zhao, D., Zhu, Y., Dong, Y., and Liu, Y. (2019). Long Non-coding RNA GAS5 Promotes Osteogenic Differentiation of Bone Marrow Mesenchymal Stem Cells by Regulating the miR-135a-5p/FOXO1 Pathway. *Mol. Cell Endocrinol.* 496, 110534. doi:10.1016/j.mce.2019.110534

Yu, L., and Liu, Y. (2019). circRNA_0016624 Could Sponge miR-98 to Regulate BMP2 Expression in Postmenopausal Osteoporosis. *Biochem. Biophysical Res. Commun.* 516, 546–550. doi:10.1016/j.bbrc.2019.06.087

Zhang, N., Chow, S. K., Leung, K. S., and Cheung, W. H. (2017). Ultrasound as a Stimulus for Musculoskeletal Disorders. *J. Orthop. Translat.* 9, 52–59. doi:10.1016/j.jot.2017.03.004

Zhao, K., Zhao, Q., Guo, Z., Chen, Z., Hu, Y., Su, J., et al. (2018). Hsa_Circ_0001275: A Potential Novel Diagnostic Biomarker for Postmenopausal Osteoporosis. *Cell Physiol. Biochem.* 46, 2508–2516. doi:10.1159/000489657

Zheng, Y., Li, X., Huang, Y., Jia, L., and Li, W. (2017). The Circular RNA Landscape of Periodontal Ligament Stem Cells during Osteogenesis. *J. Periodontol.* 88, 906–914. doi:10.1902/jop.2017.170078

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

frontiers in Genetics

BRIEF RESEARCH REPORT
published: 21 December 2021
doi: 10.3389/fgene.2021.799099

Integrative RNA-Seq and ATAC-Seq Analysis Reveals the Migration-Associated Genes Involved in Antitumor Effects of Herbal Medicine Feiyanning on Lung Cancer Cells

Chenyang Wang[1], Pengxiao Li[1], Yonglin Peng[1], Ruiqi Liu[1], Xiaoting Wu[2], Sheng Tan[1], Ming Zhang[2] and Xiaodong Zhao[1]**

[1]*Shanghai Center for Systems Biomedicine, Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Jiao Tong University, Shanghai, China,* [2]*Department of Integrated Traditional Chinese and Western Medicine, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China*

Lung cancer is one of the leading causes of cancer-associated death in the world. It is of great importance to explore new therapeutic targets. Traditional Chinese medicine formula Feiyanning has been clinically administered in China for more than a decade and raised attention due to its anticancer effect in lung cancer. However, the underlying molecular mechanisms remain to be elucidated. In the present study, we carried out cellular and molecular assays to examine the antitumor activities and understand the mechanism of the Feiyanning formula in lung cancer cells. The cellular viability of Feiyanning-treated lung cancer cells was evaluated by Cell Counting Kit-8. The effect of the Feiyanning formula on cellular migration and invasion of lung cancer cells was examined by wound healing and transwell assays. Transcriptome and chromatin accessibility analysis by RNA-seq and ATAC-seq was performed to investigate the underlying molecular mechanisms. Our results revealed that the Feiyanning formula inhibited the cellular activities of proliferation, migration, and invasion in non-small cell lung cancer cell lines A549, H1975, and 95D. Furthermore, we observed that the transcriptional activity of the migration-associated genes was downregulated upon Feiyanning formula treatment in non-small cell lung cancer cells. The chromatin accessibility of the Feiyanning-treated lung cancer genome tended to decrease, and the regulation of the cellular component movement biological process and PI3K-AKT pathway were enriched among these altered genomic regions. Taken together, the present study suggested that Feiyanning formula exerted the antitumor effects by modulating the expression and chromatin accessibility levels of migration-associated genes.

Keywords: traditional Chinese medicine, Feiyanning, non-small cell lung cancer, Migration, RNA-seq, ATAC-seq

OPEN ACCESS

Edited by:
Yuriy L. Orlov,
I.M.Sechenov First Moscow State Medical University, Russia

Reviewed by:
Guojin Huang,
Affiliated Hospital of Guilin Medical University, China
Haijian Wang,
Fudan University, China
Xiaoyu Li,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Elvira Galieva,
Novosibirsk State University, Russia

****Correspondence:***
Ming Zhang
dr_zhangming@126.com
Xiaodong Zhao
xiaodongzhao@sjtu.edu.cn

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *21 October 2021*
Accepted: *22 November 2021*
Published: *21 December 2021*

Citation:
Wang C, Li P, Peng Y, Liu R, Wu X, Tan S, Zhang M and Zhao X (2021) Integrative RNA-Seq and ATAC-Seq Analysis Reveals the Migration-Associated Genes Involved in Antitumor Effects of Herbal Medicine Feiyanning on Lung Cancer Cells. Front. Genet. 12:799099. doi: 10.3389/fgene.2021.799099

INTRODUCTION

Lung cancer is a malignancy with high morbidity and mortality around the world (Siegel et al., 2018; Cao et al., 2020). Similar to other types of cancer, lung cancer possesses many characteristics, such as rapid invasion ability (Popper, 2016). To make matters worse, the majority of lung cancer patients are already at the late stage when the initial diagnosis is performed (Gridelli et al., 2015). Lung cancer patients are routinely treated with surgery (Hoy et al., 2019), radiation therapy (Aokage et al., 2017; Gensheimer and Loo, 2017; Fitzgerald and Simone, 2020), chemotherapy (Nagasaka and Gadgeel., 2018), immunotherapy (Steven et al., 2016; Doroshow et al., 2019), targeted therapy (Hirsch et al., 2017; Duma et al., 2019), or a combination of these treatments. However, these treatments have both advantages and disadvantages. For example, radiotherapy has an effect on cancer cells; however, it simultaneously causes damage to nonmalignant cells and function of the body. Although targeted therapy does not cause damage to normal cells, this therapy is effective only for patients with mutated genes, and it inevitably induces drug tolerance after treatment for a period of time (Molina et al., 2008; Kalia, 2015; Arbour and Riely., 2019; Rapoport and Anderson., 2019). Therefore, it is important to look for alternative treatments.

During the past decades, more and more natural remedies have been used in cancer therapy (Newman and Cragg., 2014). They have few side effects and can improve life quality (Xu et al., 2014). Feiyanning (FYN), a Chinese herbal formula consisting of 11 herbs, has been clinically administered for more than a decade with benefit of prolonging patients' survival (Gong et al., 2018). Very recently, it has been reported that the FYN formula can induce apoptosis of lung cancer cells through activation of the mitochondrial pathway (Zhu et al., 2021). However, the antitumor mechanism of the FYN formula has not been completely elucidated. In recent years, we have performed a mechanistic characterization of the herbal formula. For example, we reported that Jinfukang, a Traditional Chinese medicine against lung cancer, could induce cellular apoptosis (Lu et al., 2018). Moreover, we observed that Jinfukang could improve cytotoxicity when lung cancer cells are treated with the combination of Jinfukang and cisplatin, the first-line chemotherapy drug for lung cancer (Lu et al., 2016).

In this study, we carried out cellular assays to examine the effect of the FYN formula on non-small cell lung cancer cell (NSCLC) lines and performed deep-sequencing analysis to understand the underlying mechanism. We observed that the FYN formula suppressed the cellular activities of proliferation, migration, and invasion in lung cancer cells and exerted the antitumor effects *via* the transcriptional regulation of cell migration- and death-related genes. This study suggested that the FYN formula could be a potential adjunct or even therapeutic choice for lung cancer patients and provide insight into understanding the antitumor mechanism of FYN formula.

MATERIALS AND METHODS

Preparation of FYN formula

The FYN formula consists of *Astragalus membranaceus* (30 g), *Polygonatum sibiricum* (30 g), *Cornus officinalis* (15 g), *Paris polyphylla* (9 g), *Atractylodes macrocephala* (9 g), *Polistes olivaceus* (9 g), *Salvia chinensis* (30 g), *Corium bufonis* (6 g), *Ganoderma lucidum* (15 g), *Pseudobulbus cremastrae seu pleiones* (15 g), and *Epimedii folium* (15 g). The herb medicine was obtained from the pharmacy of Shanghai Chest Hospital, Shanghai Jiao Tong University (Shanghai, China). The following components were mixed as follows: distilled deionized water was added to the herb mixture (solid–liquid ratio: 1:5) and soaked for 1 h. The mixture was kept warm for about 1 h and filtered with six layers of gauze. Then, the filtrate was immediately kept at −70°C. Finally, the lyophilized powder of the FYN formula was prepared by freeze-drying at −50°C for 72 h. The prepared FYN formula frozen powder was divided into 15–20 g/tube for the subsequent cellular assays. Mass spectrographic fingerprints for the FYN formula extract were examined for quality check. The lyophilized powder was diluted to various concentrations for the subsequent assays in the present study.

Cell culture

NSCLC cell lines A549, 95D, and H1975 were provided by the Cell Bank of Chinese Academy of Sciences (Shanghai, China) and maintained with RPMI 1640 medium (Invitrogen, Carlsbad, CA, USA). The culture medium contained 1% penicillin/streptomycin (Invitrogen, Carlsbad, USA) and 10% fetal bovine serum (FBS) (Invitrogen, Carlsbad, USA). Lung cancer cells were cultured in an incubator at 37°C with 5% CO2. *Mycoplasma* contamination was monitored through PCR to make sure that all cell lines in this study were mycoplasma-free.

Cell viability assay

The cell viability was examined by the Cell Counting Kit-8 (CCK-8, Sangon, Shanghai, China). Briefly, cells were seeded at a density of 3×10^3 cells/well and cultured in 96-well plates overnight. Lung cancer cells were treated with different concentrations of FYN formula (0, 0.25, 0.5, 1, 2, and 3 mg/ml) for 24, 48, and 72 h, respectively. After treatment, 10 μl CCK-8 solution was added into each well and incubated for 3 h. The measurement of absorbance was carried out at 450 nm with a spectrophotometric plate reader (Omega Bio-Tek, Norcross, GA, USA). Three independent experiments were carried out with five replicates in each group.

Wound healing assay

Lung cancer cells were seeded at the density of 1×10^6 cells per well into 6-well plates and incubated overnight. When lung cancer cells reached 90% confluence, they were treated with the FYN formula for 24 h. In a central position of each well, a horizontal scratch was made with a 10-μl tip. The wound closure was then observed, and the images were captured at 0 and 24 h by microscope (Nikon, Tokyo, Japan). The scratch healing degree was used to determine the migration ability. Three independent experiments were carried out.

Transwell assay

In the migration assay, 2×10^4 lung cancer cells were seeded into the upper chamber (Corning, NY, USA). For the invasion assay, we diluted Matrigel and added it to the 24-well invasion chambers and maintained overnight at 37°C. A total of 4×10^4 cells were seeded into the upper well of the chamber. In order to stimulate cell migration, medium with 15% FBS was added into the lower chamber. After incubation for 24 h, lung cancer cells were fixed for 30 min and stained with 0.1% crystal violet solution for 20 min at room temperature. The non-invaded cells at the top were removed, while the cells at the bottom were photographed in three independents 20 × fields for each well. Three independent experiments were conducted.

Preparation of strand-specific RNA-Seq library and deep sequencing

Total RNA was extracted using TRIzol (Thermo Fisher Scientific, Waltham, MA, USA), and mRNA was purified using the NEBNext Poly(A) mRNA Magnetic Isolation Module Kit (NEB, New England, Ipswich, MA, USA). A strand-specific RNA-seq library was generated with NEBNext Ultra Directional RNA Library Prep Kit (NEB, New England, USA). Briefly, mRNA was fragmented into small fragments and reversely transcribed into cDNA. Then, second-strand cDNA was generated. Next, the double-strand DNA fragments were purified with AMPure beads (Beckman Coulter, Brea, CA, USA) and ligated with adapters. The resulting ligation products were amplified and sequenced with HiSeq X (Illumina, San Diego, CA, USA). The raw sequencing data could be obtained in the EMBL database (http://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-10821. The sequencing data of the control we generated in our previous study are available in the EMBL database (http://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-7237 (Yang et al., 2019).

Preparation of ATAC-Seq library and deep sequencing

A total of 50,000 cells were resuspended for nucleus isolation. Chromatin was fragmented using Tn5 transposase (Vazyme, Nanjing, China) and followed by amplification (9 cycles). Lastly, the resulting libraries were purified with the Xygen purification kit (Xygen, USA). The distribution of library fragments was examined by the Agilent 2100 bioanalyzer. Libraries were sequenced on HiSeq X. The raw sequencing data could be obtained in the EMBL database (http://www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-11023.

Bioinformatics analysis

The raw sequencing reads generated in this study were mapped to the human genome (hg19) using TopHat v2.1.1 (Trapnell et al., 2012). For RNA-seq, differentially expressed gene analysis was carried out by comparing genes in A549 and those in A549 cells treated with FYN using cuffdiff v2.2.1. Cufflinks v2.2.1 (Trapnell et al., 2012) was used to quantify their expression. The gene expression levels were measured by fragments per kilobase of transcript per million reads. Gene ontology (GO) analysis was performed using Gene Ontology Resource (http://geneontology.org/). DAVID was used for KEGG pathway analysis (https://david.ncifcrf.gov/tools.jsp). For ATAC-seq, peaks of each sample were called with MACS2. A heatmap was generated using deepTools (v3.3.0) (Ramírez et al., 2016). To visualize the ATAC-seq signals using Integrative Genomics Viewer (IGV, v2.5.3), the bam files were transformed into bigwig (bw) files with deepTools (Robinson et al., 2011). We applied HOMER (v4.4.1) to perform differential chromatin accessibility (DA) analysis with the getDifferentialPeaks algorithm (Heinz et al., 2010). Both RNA-seq and ATAC-seq data sets were analyzed with default parameters, and the p values were corrected by conventional FDR. Notably, the promoter regions were defined as—1,000 bp/+100 bp of nearest transcription start sites. If a peak falls in the 5′UTR region, it is preferentially assigned to the 5′ UTR region rather than the promoter region.

Statistical analysis

Data were exhibited as the mean ± standard deviation (SD). The differences between each group were examined with GraphPad Prism 8.0 software. We considered the results with $^*p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$ to be significantly different.

RESULTS

FYN suppresses growth of NSCLC cells

The extract of the FYN formula was prepared as reported previously (Zhu et al., 2021). CCK-8 assay was performed to examine the effect of the FYN formula on the viability of lung cancer cell lines A549, H1975, and 95D. As shown in **Figures 1A, B**, the FYN formula significantly inhibited the viability of A549 and H1975 in a dose-dependent manner after the treatment for 24, 48, and 72 h. In addition, the FYN formula also inhibited the viability of 95D (**Figure 1C**). However, we found that A549 and H1975 were more sensitive to the FYN-induced cell viability attenuation than 95D. The half inhibitory concentration (IC50) of the FYN formula in A549, H1975, and 95D was 0.73, 1.56, and 4.68 mg/ml, respectively.

FYN inhibits NSCLC cellular migration and invasion

To further explore the anticancer effect of the FYN formula, we performed wound healing assay to examine whether FYN affects the migratory activity of lung cancer cells. We observed that the cells without treatment (the control) rapidly migrated, but the migration area of FYN-treated cells was significantly decreased and this trend was more obvious when the FYN formula concentration was increasing. After a 24-h treatment with various concentrations of the FYN formula, the migration inhibition rates of A549 cells were 66.5 ± 5.0% (0.365 mg/ml, $p < 0.05$) and 89.1 ± 3.9% (0.73 mg/ml, $p < 0.001$), respectively (**Figure 2A**). For H1975 and 95D cells, the migration inhibition rates were 81.8 ± 10.3% (0.78 mg/ml, $p < 0.01$), 89.9 ± 3.8% (1.56 mg/ml, $p < 0.001$) and

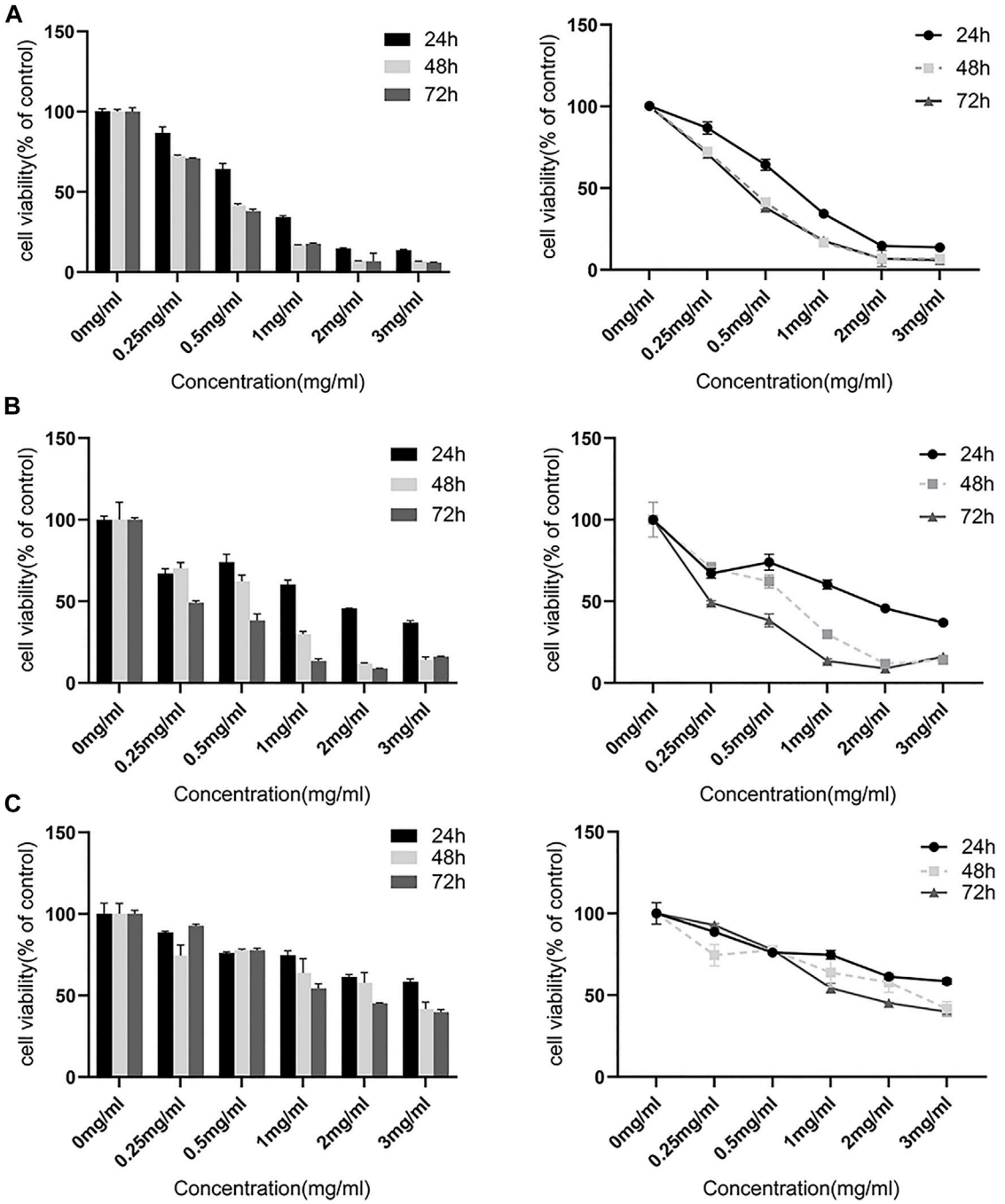


FIGURE 1 | FYN suppresses the cellular growth of lung cancer cells. The CCK-8 kit was used to detect cell viability after 24, 48, and 72 h. Data were presented as the means ± SD from three independent experiments with five replicates per experiment. **(A)**: A549, **(B)**: H1975, **(C)**: 95D.

80.8 ± 6.3% (2.34 mg/ml, $p < 0.001$), 84 ± 5.2% (4.68 mg/ml, $p < 0.001$), respectively (**Supplementary Figures S1A, B**). These results showed that the FYN formula could effectively decrease the migration of lung cancer cells examined.

We next performed transwell assay to further investigate the inhibitory effect of the FYN formula on the migration and invasion of A549, H1975, and 95D cells. We found that the numbers of A549, H1975, and 95D cells that migrated to the lower chamber declined in a concentration-dependent manner after a 24-h treatment (**Figure 2B**, **Supplementary Figure S2**). Furthermore, the density of the FYN-treated cells declined as well (**Figure 2B**, **Supplementary Figure S2**). These results further indicated that the FYN formula could obviously inhibit the migration ability of A549, H1975, and 95D cells.

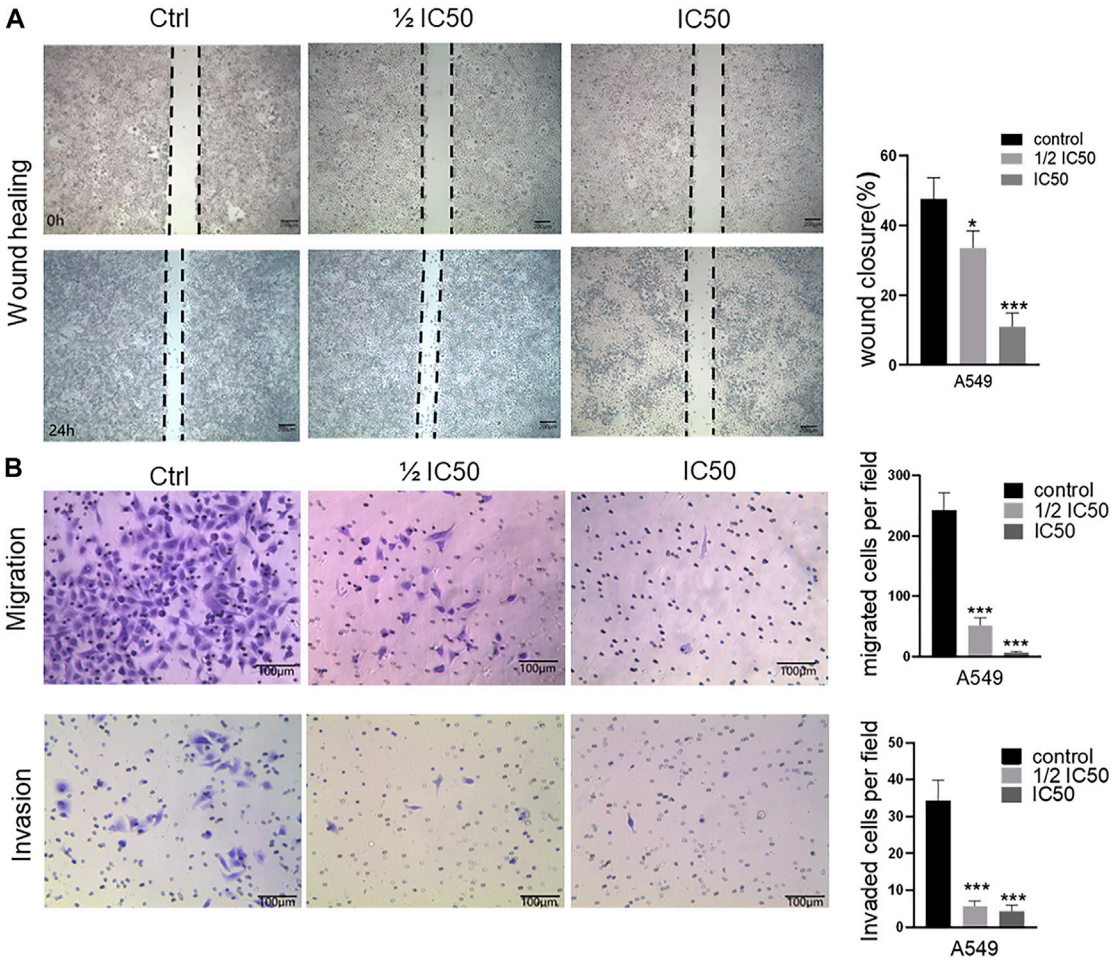


FIGURE 2 | FYN suppresses the cellular migration and invasion of A549 cells. FYN-treated A549 were used for wound healing assay (10 × field) and invasion assay (20 × field). Data were presented as the means ± SD from three independent experiments. *, **, and *** indicate significant difference compared to the control group at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. **(A)**: Wound healing assay. **(B)**: Transwell assay.

FYN-associated transcriptome analysis in treated A549 cells

Since we observed the anti-proliferation and anti-migration activities of the FYN formula, we next aimed to investigate the underlying molecular mechanisms through RNA-seq analysis. A549 cells are most sensitive to FYN among the NSCLC cells examined; we then generated RNA-seq data of FYN-treated A549 cells and set the data of A549 without FYN treatment as the control. Compared with the control, we identified 1,630 differentially expressed genes (adjusted p-values ≤ 0.05, $|\log_2\text{fold change}| \geq 1$) in FYN-treated A549; 526 genes were upregulated and 1,104 genes were downregulated (**Figure 3A**; Supplementary Table S1). GO analysis of obviously downregulated genes ($\log_2$ fold change < -4) (**Supplementary Table S2**) indicated that the GO terms of migration- and death-related biological processes were statistically enriched (**Figure 3B**). In addition, we carried out KEGG pathway analysis with the differentially expressed genes. As expected, we observed that the focal adhesion and PI3K-AKT pathways were enriched (**Figure 3C**). The transcriptional change in cell adhesion- and apoptosis-related genes was shown in **Figure 3D**. These results suggested that the FYN formula may exert anticancer effects through inhibition of cellular proliferation and migration.

Identifying altered chromatin accessible regions in FYN-treated A549 cells

It has been recognized that the open chromatin-accessible regions contain cis-regulatory elements and might modulate gene activity (Frank et al., 2015). We then performed chromatin accessibility profiling analysis in both A549 (the control) and FYN-treated A549 cells through ATAC-Seq. We identified 101,767, and 25,796 accessible regions in the control and FYN-treated A549, respectively. Specifically, we observed that the chromatin accessibility near transcription start sites (1 kb around transcription start sites) tended to decrease upon FYN treatment (**Figure 4A**, **Supplementary Figure S3A**).

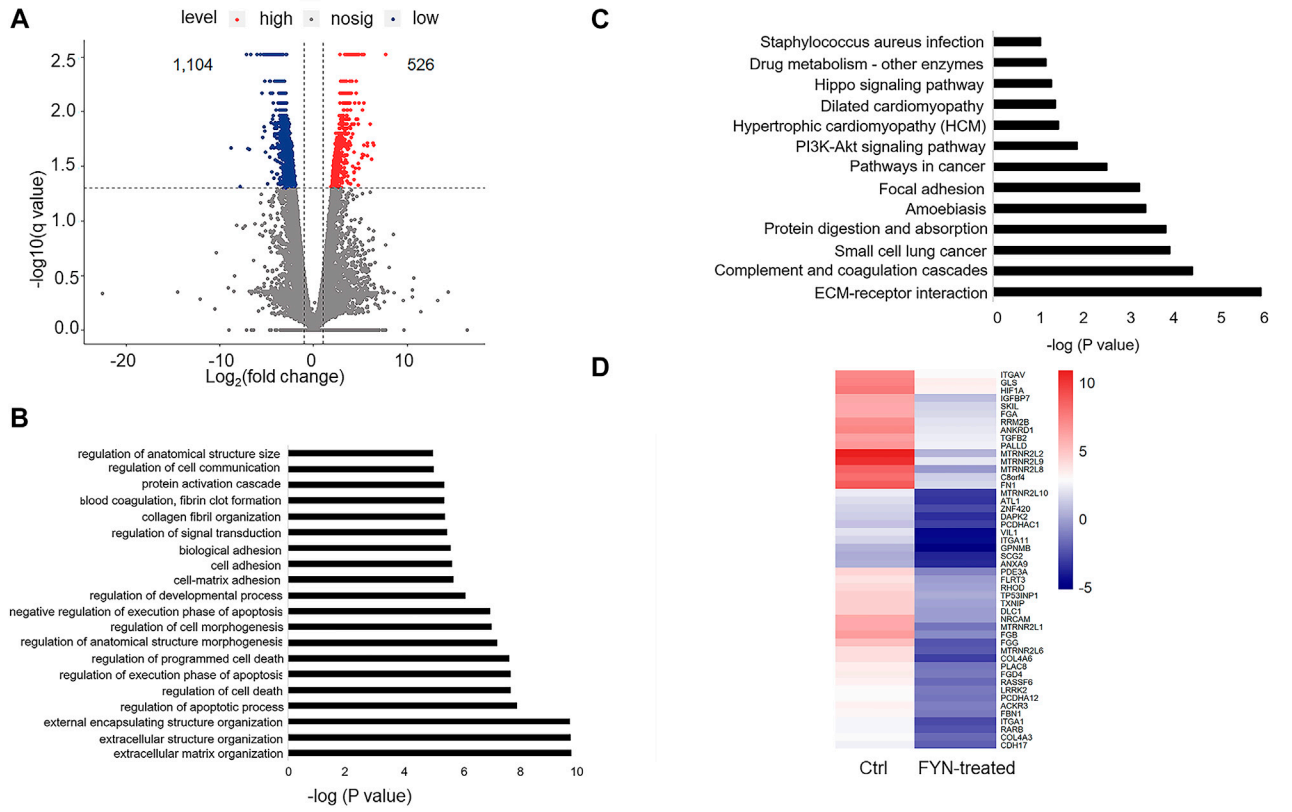


FIGURE 3 | Characterization of the FYN-induced gene expression alteration in lung cancer cells. **(A)**: Scatter plot indicated the pattern of differential gene expression in FYN-treated A549 cells and the counterpart of A549 cells without FYN treatment, including both upregulated (red) and downregulated (blue) genes. **(B)**: GO analysis of the downregulated genes in FYN-treated A549. **(C)**: KEGG analysis of the downregulated genes in A549 treated with FYN. **(D)**: Heatmap of downregulated genes related to apoptosis and cell adhesion in A549 treated with FYN.

Compared with the control, the chromatin accessibility levels of 48,290 ATAC peaks were observed to be altered ($|\log_2$ fold change $| \geq 2$) in FYN-treated A549 cells, among which 97% were downregulated and 3% were upregulated (**Figure 4B**). Most of these altered accessibility regions were located in intronic and intergenic regions (**Supplementary Figure S3B**). While detecting the relationship between RNA expression levels and chromatin accessibility levels at each genomic region, we found the correlation at the gene promoter in FYN-treated A549 was relatively large (**Figure 4C**, **Supplementary Figure S4**).

Then, the obviously downregulated peaks ($\log_2$ fold change < -5) were annotated to the nearest gene (**Supplementary Table S3**) and we performed GO and KEGG analyses. We found that GO item regulation of cellular component movement (**Figure 4D**) and PI3K-AKT signaling pathway (**Figure 4E**) were enriched, which is consistent with the observation from RNA-seq analysis. For example, in FYN-treated A549 cells we found the decreased chromatin accessibility and lower expression level at the *COL4A6* (**Figure 4F**), a member of the PI3K-AKT signaling pathway that was reported to be involved in cell metastasis (Ma et al., 2020). These findings suggested that the migration-associated genes were involved in the antitumor effect of FYN formula.

DISCUSSION

There is mounting evidence that herbal medicines play increasingly important roles in cancer treatment, as adjunctive treatments or even therapeutic agents (Liu et al., 2012; Jiang et al., 2016). The FYN formula is composed of 11 herbs, which has been found to have anticancer efficacy in lung cancer (Xu et al., 2011). In this study, we observed that the FYN formula exerted antitumor effects by inhibiting the cellular proliferation, migration, and invasion of A549, H1975, and 95D cells. Furthermore, we found that FYN treatment suppressed the expression of migration- and focal adhesion-associated genes and altered chromatin accessibility. In particular, among the genes with suppressed expression levels and decreased chromatin accessible regions, we found that the PI3K-AKT pathway was enriched (**Figures 3C**, **4E**). The PI3K-AKT pathway has been reported to be activated in NSCLC, and some small compound inhibitors of PI3K and AKT were developed for clinical trials in NSCLC patients (Tan, 2020). In our study, we found that the FYN formula also inhibited the PI3K-AKT pathway, which plays a similar role as those small compound inhibitors do. Collectively, these findings may provide new evidence to support the antitumor potential of the FYN formula as an adjunctive treatment for NSCLC.

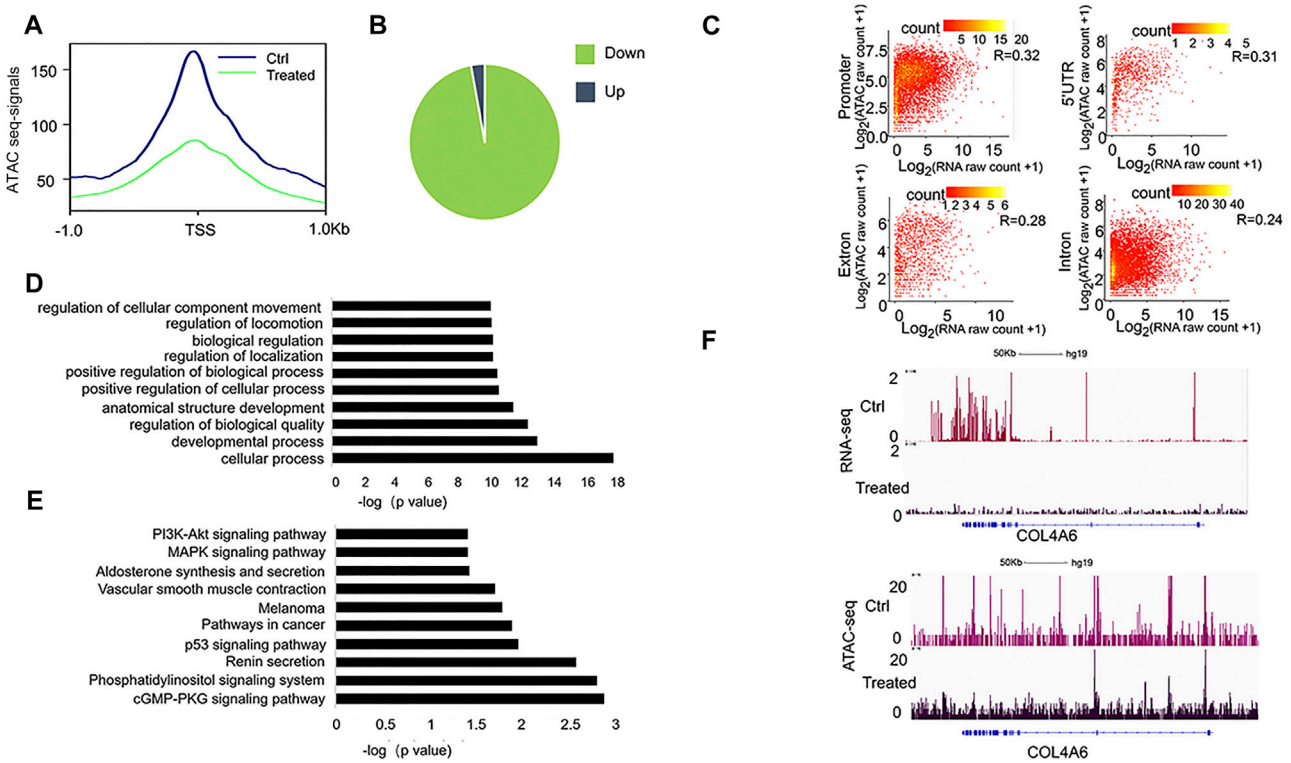


FIGURE 4 | Characterization of genome-wide chromatin accessibility in FYN-treated A549 cells. **(A)**: Line plot shows the decreased chromatin accessibility in FYN-treated A549 cells around the TSSs of the nearest genes. **(B)**: The pie chart shows the chromatin accessibility levels after FYN treatment. **(C)**: Point density plots indicating the correlation between ATAC-seq total tags and RNA-seq expression levels at the promoter, 5′UTR, exon, and intron regions in FYN-treated A549. **(D)**: GO analysis of differentially accessible peaks. **(E)**: KEGG of differentially accessible peaks. **(F)**: IGV browser snapshots at the *COL4A6* locus showing a lower expression level and decreased chromatin accessibility.

We were aware of the limitations of this research. Firstly, the biological consequence of FYN-induced expression alteration of cell growth- and migration-related genes needs further investigation. Secondly, in this study we carried out cellular assays and differential gene expression analysis. It would be more informative to validate the present findings in the mouse xenograft of human lung cancer.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

XZ and ST: conception, supervision, and manuscript revising. XW, XZ, and MZ: resources. CW, RL, PL, and YP: experiment conduction and data analysis. XZ and MZ: project administration and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by Shanghai Jiao Tong University Scientific and Technological Innovation (Fund No. 2019TPA09), Interdisciplinary Program of Shanghai Jiao Tong University (Grant No. YG2019ZDA25), and Key Project of Innovation of Shanghai Education Committee (Grant No. 2017-01-07-00-10-E00064).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.799099/full#supplementary-material

REFERENCES

Aokage, K., Yoshida, J., Hishida, T., Tsuboi, M., Saji, H., Okada, M., et al. (2017). Limited Resection for Early-Stage Non-small Cell Lung Cancer as Function-Preserving Radical Surgery: a Review. *Jpn. J. Clin. Oncol.* 47, 7–11. doi:10.1093/jjco/hyw148

Arbour, K. C., and Riely, G. J. (2019). Systemic Therapy for Locally Advanced and Metastatic Non-small Cell Lung Cancer. *JAMA* 322, 764–774. doi:10.1001/jama.2019.11058

Cao, M., Li, H., Sun, D., and Chen, W. (2020). Cancer burden of Major Cancers in China: A Need for Sustainable Actions. *Cancer Commun.* 40, 205–210. doi:10.1002/cac2.12025

Doroshow, D. B., Sanmamed, M. F., Hastings, K., Politi, K., Rimm, D. L., Chen, L., et al. (2019). Immunotherapy in Non-small Cell Lung Cancer: Facts and Hopes. *Clin. Cancer Res.* 25, 4592–4602. doi:10.1158/1078-0432.ccr-18-1538

Duma, N., Santana-Davila, R., and Molina, J. R. (2019). Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clinic Proc.* 94, 1623–1640. doi:10.1016/j.mayocp.2019.01.013

Fitzgerald, K., and Simone, C. B., 2nd (2020). Combining Immunotherapy with Radiation Therapy in Non-small Cell Lung Cancer. *Thorac. Surg. Clin.* 30, 221–239. doi:10.1016/j.thorsurg.2020.01.002

Frank, C. L., Liu, F., Wijayatunge, R., Song, L., Biegler, M. T., Yang, M. G., et al. (2015). Regulation of Chromatin Accessibility and Zic Binding at Enhancers in the Developing Cerebellum. *Nat. Neurosci.* 18, 647–656. doi:10.1038/nn.3995

Gensheimer, M. F., and Loo, B. W., Jr (2017). Optimal Radiation Therapy for Small Cell Lung Cancer. *Curr. Treat. Options. Oncol.* 18, 21. doi:10.1007/s11864-017-0467-z

Gong, Y., Xu, Z., Jin, C., Deng, H., Wang, Z., Zhou, W., et al. (2018). Treatment of Advanced Non-small-cell Lung Cancer with Qi-Nourishing Essence-Replenishing Chinese Herbal Medicine Combined with Chemotherapy. *Biol. Proced. Online* 20, 9. doi:10.1186/s12575-018-0074-9

Gridelli, C., Rossi, A., Carbone, D. P., Guarize, J., Karachaliou, N., Mok, T., et al. (2015). Non-small-cell Lung Cancer. *Nat. Rev. Dis. Primers* 1 (1), 15009. doi:10.1038/nrdp.2015.9

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cel* 38, 576–589. doi:10.1016/j.molcel.2010.05.004

Hirsch, F. R., Scagliotti, G. V., Mulshine, J. L., Kwon, R., Curran, W. J., Jr, Wu, Y.-L., et al. (2017). Lung Cancer: Current Therapies and New Targeted Treatments. *The Lancet* 389, 299–311. doi:10.1016/s0140-6736(16)30958-8

Hoy, H., Lynch, T., and Beck, M. (2019). Surgical Treatment of Lung Cancer. *Crit. Care Nurs. Clin. North America* 31, 303–313. doi:10.1016/j.cnc.2019.05.002

Jiang, Y., Liu, L.-S., Shen, L.-P., Han, Z.-F., Jian, H., Liu, J.-X., et al. (2016). Traditional Chinese Medicine Treatment as Maintenance Therapy in Advanced Non-small-cell Lung Cancer: A Randomized Controlled Trial. *Complement. Therapies Med.* 24, 55–62. doi:10.1016/j.ctim.2015.12.006

Kalia, M. (2015). Biomarkers for Personalized Oncology: Recent Advances and Future Challenges. *Metabolism* 64 (3 Suppl. 1), S16–S21. doi:10.1016/j.metabol.2014.10.027

Liu, T. G., Xiong, S. Q., Yan, Y., Zhu, H., and Yi, C. (2012). Use of Chinese Herb Medicine in Cancer Patients: a Survey in Southwestern china. *Evid. Based Complement. Alternat Med.* 2012, 769042. doi:10.1155/2012/769042

Lu, J., Chen, J., Kang, Y., Wu, J., Shi, H., Fu, Y., et al. (2018). Jinfukang Induces Cellular Apoptosis through Activation of Fas and DR4 in A549 Cells. *Oncol. Lett.* 16, 4343–4352. doi:10.3892/ol.2018.9149

Lu, J., Chen, J., Xu, N., Wu, J., Kang, Y., Shen, T., et al. (2016). Activation of AIFM2 Enhances Apoptosis of Human Lung Cancer Cells Undergoing Toxicological Stress. *Toxicol. Lett.* 258, 227–236. doi:10.1016/j.toxlet.2016.07.002

Ma, J.-B., Bai, J.-Y., Zhang, H.-B., Gu, L., He, D., and Guo, P. (2020). Downregulation of Collagen COL4A6 Is Associated with Prostate Cancer Progression and Metastasis. *Genet. Test. Mol. Biomarkers* 24, 399–408. doi:10.1089/gtmb.2020.0009

Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., and Adjei, A. A. (2008). Non-small Cell Lung Cancer: Epidemiology, Risk Factors, Treatment, and Survivorship. *Mayo Clinic Proc.* 83, 584–594. doi:10.1016/s0025-6196(11)60735-0

Nagasaka, M., and Gadgeel, S. M. (2018). Role of Chemotherapy and Targeted Therapy in Early-Stage Non-small Cell Lung Cancer. *Expert Rev. Anticancer Ther.* 18, 63–70. doi:10.1080/14737140.2018.1409624

Newman, D. J., and Cragg, G. M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661. doi:10.1021/acs.jnatprod.5b01055

Popper, H. H. (2016). Progression and Metastasis of Lung Cancer. *Cancer Metastasis Rev.* 35, 75–91. doi:10.1007/s10555-016-9618-0

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., et al. (2016). deepTools2: a Next Generation Web Server for Deep-Sequencing Data Analysis. *Nucleic Acids Res.* 44, W160-W165. doi:10.1093/nar/gkw257

Rapoport, B., and Anderson, R. (2019). Realizing the Clinical Potential of Immunogenic Cell Death in Cancer Chemotherapy and Radiotherapy. *Ijms* 20, 959. doi:10.3390/ijms20040959

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative Genomics Viewer. *Nat. Biotechnol.* 29, 24–26. doi:10.1038/nbt.1754

Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer Statistics, 2018. *CA: A Cancer J. Clinicians* 68, 7–30. doi:10.3322/caac.21442

Steven, A., Fisher, S. A., and Robinson, B. W. (2016). Immunotherapy for Lung Cancer. *Respirology* 21, 821–833. doi:10.1111/resp.12789

Tan, A. C. (2020). Targeting the PI3K/Akt/mTOR Pathway in Non-small Cell Lung Cancer (NSCLC). *Thorac. Cancer* 11, 511–518. doi:10.1111/1759-7714.13328

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi:10.1038/nprot.2012.016

Xu, W., Yang, G., Xu, Y., Zhang, Q., Fu, Q., Yu, J., et al. (20142014). The Possibility of Traditional Chinese Medicine as Maintenance Therapy for Advanced Nonsmall Cell Lung Cancer. *Evid. Based Complement. Alternat Med.* 2014, 278917. doi:10.1155/2014/278917

Xu, Z. Y., Jin, C. J., Zhou, C. C., Wang, Z. Q., Zhou, W. D., Deng, H. B., et al. (2011). Treatment of Advanced Non-small-cell Lung Cancer with Chinese Herbal Medicine by Stages Combined with Chemotherapy. *J. Cancer Res. Clin. Oncol.* 137, 1117–1122. doi:10.1007/s00432-011-0975-3

Yang, W., Kang, Y., Zhao, Q., Bi, L., Jiao, L., Gu, Y., et al. (2019). Herbal Formula Yangyinjiedu Induces Lung Cancer Cell Apoptosis via Activation of Early Growth Response 1. *J. Cel Mol Med* 23, 6193–6202. doi:10.1111/jcmm.14501

Zhu, L.-M., Shi, H.-X., Sugimoto, M., Bandow, K., Sakagami, H., Amano, S., et al. (2021). Feiyanning Formula Induces Apoptosis of Lung Adenocarcinoma Cells by Activating the Mitochondrial Pathway. *Front. Oncol.* 11, 690878. doi:10.3389/fonc.2021.690878

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 24 January 2022
doi: 10.3389/fgene.2021.784135

A Syndrome of Variable Allergy, Short Stature, and Fatty Liver

Jing Qiao[1], Yue Chen[1], Ying Lu[2], Tiejun Wang[3], Xiaoli Li[1], Wei Qin[1], Aifen Li[4] and Guangquan Chen[5]*

[1]Department of Pediatrics, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China, [2]Department of Clinical Laboratory, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China, [3]Department of Pharmacy, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China, [4]Department of Pediatrics, Jian Hospital of Shanghai East Hospital, Jian, China, [5]Fetal Medicine Unit & Prenatal Diagnosis Center, Shanghai First Maternity and Infant Hospital, Tongji University, Shanghai, China

SLC22A18 (solute carrier family 22 member 18) is an imprinted gene, but its role in growth and development is not clear. In the present study, we recorded the clinical information of six male patients of six unrelated families. Real-time quantitative PCR, Sanger sequencing, and DNA methylation sequencing were performed in these patients. The results suggested that the patients with the clinical characteristics of allergic allergy, short stature, and fatty liver had a lower expression of *SLC22A18*. One novel variant (chr11: 2899732 delA) with clinical significance was found in the core promoter region of the patients. Overall, this study found a syndrome associated with *SLC22A18*.

Keywords: *SLC22A18*, imprinted gene, allergy, short stature, fatty liver

OPEN ACCESS

Edited by:
Hua Li,
Shanghai Jiao Tong University, China

Reviewed by:
Guodong Ding,
Shanghai Children's Hospital, China
Wei Li,
Naval Medical University, China

****Correspondence:***
Jing Qiao
qiaojing@tongji.edu.cn

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *27 September 2021*
Accepted: *26 November 2021*
Published: *24 January 2022*

Citation:
Qiao J, Chen Y, Lu Y, Wang T, Li X, Qin W, Li A and Chen G (2022) A Syndrome of Variable Allergy, Short Stature, and Fatty Liver. Front. Genet. 12:784135. doi: 10.3389/fgene.2021.784135

INTRODUCTION

The prevalence of allergic diseases has been increasing worldwide over the past 60 years, affecting about 30% of the global population (Palomares et al., 2017). A phenomenon known as "allergic march" had been firstly described by Fouchard in 1973. It is a process from infant eczema to food allergy, asthma, and rhinitis resulting in poor quality of life in childhood (Sohi and Warner, 2008).

In clinical practice, a common triad, including variable allergies, short stature, and fatty liver, has not been reported as a syndrome up to now. Previous studies on allergic diseases mainly focused on the immunogenic origin of allergic diseases, the clinical significance of "health hypothesis," and the impact of maternal and infant nutrition on allergic epidemics and paid little attention to the role of human imprinted genes.

SLC22A18 is an imprinted gene, which is involved in tumor suppression and lipid accumulation. Diseases associated with *SLC22A18* include lung cancer and breast cancer (Dao et al., 1998; Peters, 2014; Ito et al., 2019), but its role in childhood diseases is not clear.

In this study, we describe six male patients from six unrelated families with a triad symptom of progressive postnatal slow growth, allergies, and fatty liver. After real-time quantitate PCR (RT-qPCR), Sanger sequencing, and DNA methylation sequencing analysis, we showed that all the patients had a lower expression of *SLC22A18* that resulted from abnormal methylation-hampered promoter function. These cases and analysis indicate a syndrome associated with *SLC22A18*.

METHODS

Cases

From Nov 2013 to Aug 2020, six male patients from six unrelated families who were admitted to the pediatric endocrinology clinic presented a triad symptom of progressive postnatal slow growth,

allergies, and fatty liver. After reviewing the patients' family history, we found that the patients' grandfathers or fathers also had similar growth experience compared with the patients. Subsequently, the medical history, physical findings, and the results of hematology, biochemistry, radiology, type B ultrasonic test, and molecular biology tests were studied. B-ultrasound examination found that the patients' liver had infiltration, suggesting that the patient had fatty liver. All laboratory procedures for clinical samples have been reported in advance. Blood, feces, and urine samples were taken; plasma was separated in the EDTA bottles; and serum was separated in the clotting blood bottles. This study was approved by the Institutional Review Board and Ethics Committee. Written consent from all patients was collected.

Real-Time Quantitate PCR

Whole blood of 10 patients (IDs: 19010101, 19010102, 19010104, 190101010, 19010111, LXY, DTY, LC, OYZY, and GJX) and 10 healthy controls were first processed with Red cell lysis buffer (Sangon Biotech, Shanghai, China) and then treated with TRIzol (Invitrogen, Carlsbad, CA, USA) to extract total RNA. Reverse transcription was performed with HiScript First Strand cDNA Synthesis Kit (Vazyme, Nanjing, China) to obtain cDNAs. Then qPCR was performed on Bio-Rad CFX96 (Bio-Rad Laboratories Inc., Hercules, CA, USA) with AceQ qPCR SYBR Green Master Mix (without ROX) (Vazyme, Nanjing, China) according to the manufacturers' protocols. *ACTB* gene was used as the reference, and the primer sequences are listed in **Supplementary Table S1**. Cycling conditions were as follows: 95°C for 5 min, followed by 40 cycles of 95°C for 10 s and 60°C for 25 s.

DNA Methylation Sequencing and Data Analysis

Genomic DNA was extracted from whole blood using TIANamp Blood DNA Kit (Tiangen Biotech, Beijing, China), which was further treated with EpiTect Fast DNA Bisulfite Kit (QIAGEN, Hilden, Germany) for bisulfite conversion. The converted DNA was PCR-amplified with primer sequences designed to cover the CpGs in two promoter regions denoted as "Promoter 1" and "Promoter 2" (**Supplementary Table S2**): each region included near 1,000 bp centering around the transcription start sites (TSSs) of SLC22A18; the TSS annotation was based on RefSeq release 109. The PCR products were gel- and column-purified and used for DNA library preparation. The library was prepared with KAPA HTP Library Preparation Kit (KAPA Biosystems, Wilmington, MA, USA) according to the manufacturer's protocol. The library was further amplified for 10 cycles, which was then subjected to deep sequencing on the Illumina HiSeq platform with 2 × 150 as the sequencing mode.

Raw reads were filtered to obtain high-quality clean reads by removing sequencing adapters and low-quality reads using Trim Galore (v0.5.0) with parameters--paired--rrbs--illumina--fastqc (https://github.com/FelixKrueger/TrimGalore) (FastQC, 2010; Martin, 2011). The clean reads were mapped to human genome (hg38) using the Bismark (v0.7.0) software (Krueger and Andrews, 2011). The methylation percentages for the CpG sites were calculated by the Bismark methylation extractor script from Bismark. Differentially methylated CpGs (DMCs) were identified using methylKit with the *q*-value cutoff set to 0.01 (Akalin et al., 2012). Differentially methylated regions (DMRs) between patients and healthy controls were identified within the two promoter regions using methylKit, which had a *q*-value of less than 0.01 and at least one DMC inside.

Sanger Sequencing

Genomic DNA was extracted from the whole blood of nine patients (IDs: 16, 17, 20, 170609, 17071201, 19010101, 19010106, 19010109, and TANG) with the same method described above. Two pairs of PCR primers were used to amplify the promoter regions of SLC22A18 (**Supplementary Table S3**), yielding close to 500 bp flanking the TSSs. PCRs were performed in a 50-μl reaction containing 10 μM of each primer, 100 ng of genomic DNA, and 25 μl of 2xFtaq PCR MasterMix (Zoman Biotechnology, Beijing, China). Cycling conditions were as follows: 95°C for 5 min, followed by 40 cycles of 95°C for 15 s, 60°C for 15 s, and 72°C for 50 s. The PCR products were gel- and column-purified and then sequenced with ABI 3730XL (Applied Biosystems, Foster City, CA, USA).

RESULTS

Clinical Features of the Six Cases

The patient in case 1 was short and light, and his father had marked central obesity (**Figure 1A**). In case 2, the patient had short stature, was lightweight, and has small hands and teethed at a normal age, and his father had marked central obesity (**Figure 1B**). The patient of case 3 is the brother of boy–girl twins. He was thinner and shorter than his twin sister before 12 years old (left picture). However, after 1.5 years of recombinant human growth hormone (rhGH) treatment, he is 9 cm taller than his twin sister (middle picture) now. His father suffered from central obesity (**Figure 1C**). The patient of case 4 was of short stature and lightweight and has small hands. He teethed at a normal age, and his father had marked central obesity (**Figure 1D**). The patient of case 5 has short stature without spinal scoliosis, was lightweight, and has small hands, and his father had marked central obesity (**Figure 1E**). The patient in case 6 was short, with a body mass index (BMI) of 18.0, and had fatty liver, and his father had central obesity (**Figure 1F**).

Case 1 The initial dose of rhGH is 2-3 IU/d over a period of 18 months and he received total 1248 IU of rhGH. During this period, his height and weight increased by 12.9 cm (to 121.0 cm) and 6.8 kg (to 23.0 kg). Then, he suspended injection rhGH since his height became normal compared to his peers, after approximately 10 months, his height and weight was still 121.0 cm and 23.0 kg, without any appropriate increase. Then started using rhGH, 4 IU/d, 6 days/week. Up to date, when he was 9y9m, both of his height and weight were normal (137.9 cm, 30.3 kg). At this time, the dose of rhGH is 5 IU/d, 6 days/week.

Case 2 When he was 3y10m years old, the injections were started and lasted for approximately 1.5 years, by a frequency of rhGH 2 IU/d, 6 days/week. Afterwards, he had a remarkable height increase which is from 95.6 cm to 112.5 cm. Sometimes he suspended for 3 months because the great treatment effects. Last revisited at Jan, 1, 2019, his height was 128.5 cm (normal), and weight was 25.3 kg.

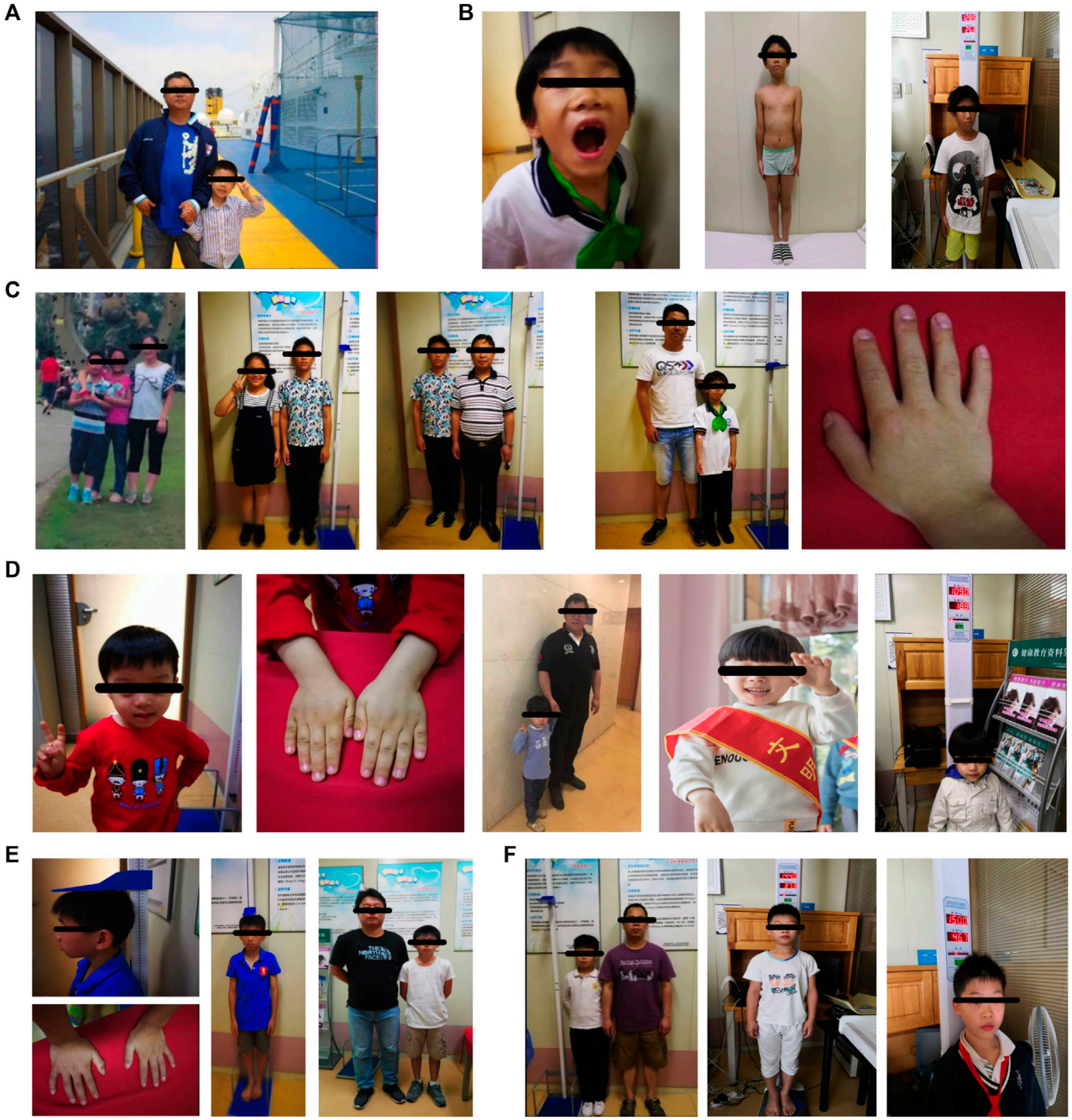


FIGURE 1 | Clinical features of the six cases' profiling.

Case 3 When he was 12.3 years old, given him VitD3 supplements and treated with rhGH 6 IU/d. The tratments were continued for approximately 3 years (from Nov. 2015 to July. 2018). Last revisited was at 14.8 years old, and his height was 171.0 cm and weight was 67.5 kg.

Case 4 To data, his height was 101.5 cm (-3SD) and weight was 17.0 kg (-1SD), when he was five years old. Injection of rhGH with 2 IU/d was recommended. After 4 months, his height was 109.0 cm and weight was 17.0 kg, which means effective treatments, however, he still had cough and lean body mass.

Case 5 He received 5 IU/d rhGH, over a period of six months, with height increased from 137.5 cm to 145.0 cm, and weight was 34.7 kg (normal height increases by 3.0 cm in six months). Treatment was interrupted for the following six months, resulting in no further growth. We suggested the patient visiting E.N.T. department for his sleep problems and traditional Chinese medicine for his poor of appetite. Now, he is 158.0 cm and 46.0 kg, and his appetite has improved.

Case 6 When he was 11.5 years old, he started using rhGH (4 IU/d). Well, at that time of Aug 23, 2018 his height was 144.0 cm

TABLE 1 | Summary of main clinical features and laboratory results of six cases with SLC22A18-associated syndrome.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Age (at first examination) (year–month)	5 years + 3 months	3 years + 7 months	12 years	4 years + 2 months	11 years + 11 months	11 years + 2 months
Gender	Male	Male	Male	Male	Male	Male
Patient's height (cm)	107.1	94.5	143.8	97.5	137.5	140
Patient's weight (kg)	16.3	13.4	44.5	14.4	29.5	34.7
Patient's BMI (kg/m^2)	14.17	15.01	21.52	15.15	15.6	17.7
Birth' history Gestation (weeks)	40w+2	40w–4	40w–3	39w+6	41 + 1	40w+4
Birth height (cm)	50	49	50	50	45	50
Birth weight (kg)	3.4	3.25	3.1	3.3	3.3	3.8
Father's height (cm)	160	167	169.4	169.6	167.5	161.5
Father's weight (kg)	74.3	70	91.7	108	83	84.5
Father's BMI (kg/m^2)	29.02	25.1	31.96	37.55	29.58	32.4
Father's chronic medical illness	Fatty liver central obesity	Fatty liver	Fatty liver central obesity chronic	Fatty liver central obesity chronic	Fatty liver central obesity	Fatty liver central obesity chronic
Mother's height (cm)	152.7	155	152.8	154	154	158
Mother's weight (kg)	47.2	47	53.9	54	53	50
Mother's BMI (kg/m^2)	20.2	19.6	23.1	22.8	22.3	20
Symptoms and signs						
Low growth velocity	+	+	+	+	+	+
Short stature	+	+	+	+	+	+
low weight (low body mass)	+	+	Normal	+	+	Normal
Snoring	+	+	+	+	+	+
Adenoid hypertrophy	+	+	+	+	+	+
Allergic rhinitis	+	+	+	+	+	–
Asthma	+	–	–	+	–	+
IGF-1 (μg/ml)	354 (↓)	139	391 (↓)	240	178 (↓)	113 (↓)
IGFBP-3 (μg/ml)	5	4.79	5.7	4.1	4.8	4.4
$VitD_3$ (ng/ml) (normal range ≥30)	28.3 (↓)	35.2	23.9 (↓)	20.9 (↓)	14.9 (↓)	22.9 (↓)
Peak of GH (μg/L) (normal range >10)	8.4 (↓)	3.98 (↓)	2.4 (↓)	7.61 (↓)	19.3 (↑)	6.74 (↓)
Hemoglobin (normal range 120.0–140.0 g/L)	126	118 (↓)	152	127	145	112 (↓)
Neutrophils (%) (normal range 40.0–75.0)	49	32.1	74.5	54.6	59	63.7
Lymphocytes (%) (normal range 20.0–50.0)	36.5	54.9	13.5	30	32.5	25.9
Monocytes (%) (normal range 1.0–8.0)	8.4 (↑)	8.4 (↑)	8.4 (↑)	12.6 (↑)	4.7	10.1 (↑)
Eosinophils (%) (normal range 0.4–8.0)	5.2	5.2	3.1	2.5	2.9	0.0 (↓)
Determination of serum allergens	Pollen	Mite		Pollen, egg	Pollen	Mite, egg
TG (mmol/L) (normal range 0–5.2)	4.2	4.5	4.2	4.6	4.2	4.26
TC (mmol/L) (normal range 0–2.26)	0.7	0.6	0.9	1.25	0.5	1.46
HDL-C (mmol/L) (normal range ≥1.04)	2.2	1.8	1.5	1.3	1.9	1.72
LDL-C (mmol/L) (normal range ≤3.34)	1.9	2.6	2.6	3.04	2.37	2.4
APOA (g/L) (normal range 1.04–2.02)	1.76	1.97	1.5	1.3	1.54	1.68
APOB (g/L) (normal range 0.66–1.33)	0.5 (↓)	0.8	0.8	0.9	0.69	0.76
APOE (mg/L) (normal range 27–45)	55.9 (↑)	35.3	50.4 (↑)	62.0 (↑)	39	54.0 (↑)
FFAs (mmol/L) (normal range 0.1–0.6)	0.95 (↑)	1.15 (↑)	0.86 (↑)	0.86 (↑)	0.67 (↑)	0.94 (↑)
Glucose (mmo1/L) (normal range 4.11–6.05)	4.8	4.5	5.4	4.6	4.3	4.9
HbA1c (%) (normal range 4–6)	4	4.1	5	4.3	4.2	5
Insulin (%) (normal range 2.6–24.9)	3.7	4.4	5.5	3.7	4.5	5.1

(Continued on following page)

TABLE 1 | (*Continued*) Summary of main clinical features and laboratory results of six cases with SLC22A18-associated syndrome.

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Liver and renal function		Normal	Normal	Normal	Normal	Normal	Normal
Fatty liver (ultrasonic B)		–	–	+	–	–	+
Pituitary MRI		Normal	Normal	Normal	Normal	Normal	Normal
Chronological age (CA) (year)		5.3 years	7 years	14 years	4 years	12 years	10.5 years
Bone age (BA) (year)		2.5 years	4.5 years	13.5 years	1.5 years+	9 years–	8 years
X-ray of BA and CA		BA < CA	BA < CA	BA = CA	BA < CA	BA < CA	BA < CA
Diagnosis							
Endocrinologist		GHD	GHD	Short stature	GHD	ISS Puberty state	GHD Puberty state
Otolaryngologist		Adenoid	Adenoid	Adenoid	Adenoid	Adenoid	Adenoid
Pulmonary physicians					Asthma		
Gastroenterology				Fatty liver			Fatty liver
Treatment	rhGH	+	+	+	+	+	+
	Adenoidectomy	+	+	–	+	–	+
	Antiallergic therapy	+	–	–	+	–	+
	Vitamin D_3	+	+	+	+	+	+

Note. BMI, body mass index; TG, triglyceride; TC, total cholesterol; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; FFA, free fatty acid; HbA1c, glycated hemoglobin; GHD, growth hormone deficiency; rhGH, recombinant human growth hormone.

and weight was 37.3 kg. Suggested him using rhGH (5 IU/d) and the treatment effects were very effective for height.

All cases were male with normal birth height and weight. When these patients visited our department, they had low body weight and slow growth, similar to idiopathic short stature (ISS).

All cases had hypertrophy of adenoid and tract allergy of the upper respiratory tract. Besides, we found that the grandfathers or fathers of the six patients also showed nearly the same combination of short stature, allergic march, and fatty livers during their puberty. Growth hormone deficiency (GHD) was found in five cases, except in case 5. Patients had decreased vitamin D_3 ($VitD_3$) and insulin-like growth factor-1 (IGF-1) and increased free fatty acid (FFA). MRI of the pituitary showed no abnormalities, and the intelligence and sexual development of these patients were normal (**Table 1**). These patients were diagnosed with GHD or ISS, along with other diagnoses such as an allergic reaction of the upper respiratory tract, asthma, and fatty liver (**Table 1**). After $VitD_3$ supplementation and rhGH treatment, their height increased by >7 cm/year (**Figure 2**).

Quantification of *SLC22A18* Expression Level

RT-qPCR was used to measure the expression level of *SLC22A18* in both patients and healthy controls. A significant difference in expression was detected between patients and healthy controls (**Figure 3A**). On average, the expression of *SLC22A18* in the healthy control group was 1.82 times higher than that in the patient group.

Methylation Analysis and Variant Detection in Promoter Regions of *SLC22A18*

Aberrantly, DNA methylation is one of the most possible reasons to drive the expression of *SLC22A18* to become abnormal. Here, we quantified the promoter methylation levels of three patients (IDs: 19010102, 19010110, and LC) and three healthy controls by deep sequencing. Deep sequencing generates 2.7 to 4.8 million reads for each individual, leading to an ultra-high coverage (>50,000× on average) for two specific promoter regions (denoted as "Promoter 1" and "Promoter 2"), each of which included nearly 1,000 bp centering around the TSSs of *SLC22A18* (**Supplementary Table S2**). Using methylKit, DMRs were identified in Promoter 1: 1) the core promoter was the most differentially methylated between patients and healthy controls with 21% methylation percentage change (p-value <0.001); 2) the upstream 500 bp of the TSS had a marked elevated methylation percentage (15%, p-value < 0.001); and 3) the downstream 500 bp of the TSS had a relatively small increase in methylation (6% change, p-value <0.001). On the contrary, the majority of CpG sites in Promoter 2 were nearly 100% methylated, and thus no noted change in methylation level was observed in Promoter 2 (**Figure 3C**). To identify variants that possibly lead to the low expression of *SLC22A18* in patients, Sanger sequencing was performed for the two promoter regions described above. Genomic DNA was extracted from the whole blood of nine patients (IDs: 16, 17, 20, 170,609, 17,071,201, 19010101, 19010106, 19010109, and TANG), yielding close to 500 bp flanking the TSSs that were subjected to Sanger sequencing. The sequencing results were aligned to the human reference genome (GRCh38) and visualized by novoSNP (Weckx et al., 2005). A total of seven variants were identified and, as expected, most of them (6) could be found in dbSNP (i.e., rs365605, rs5789280, rs538924456, rs397933484, rs366696, and rs367035) without clinical significance (Sherry et al., 2001; Landrum et al., 2014). Only one variant (chr11: 2899732 delA), located in the core promoter region, has not been reported before (**Figure 3B**). To further check if this novel variant could be related to the expression of *SLC22A18*, the 100-bp DNA sequence flanking this variant was submitted to JASPAR database (https://jaspar.genereg.net/) to detect possible binding sites of

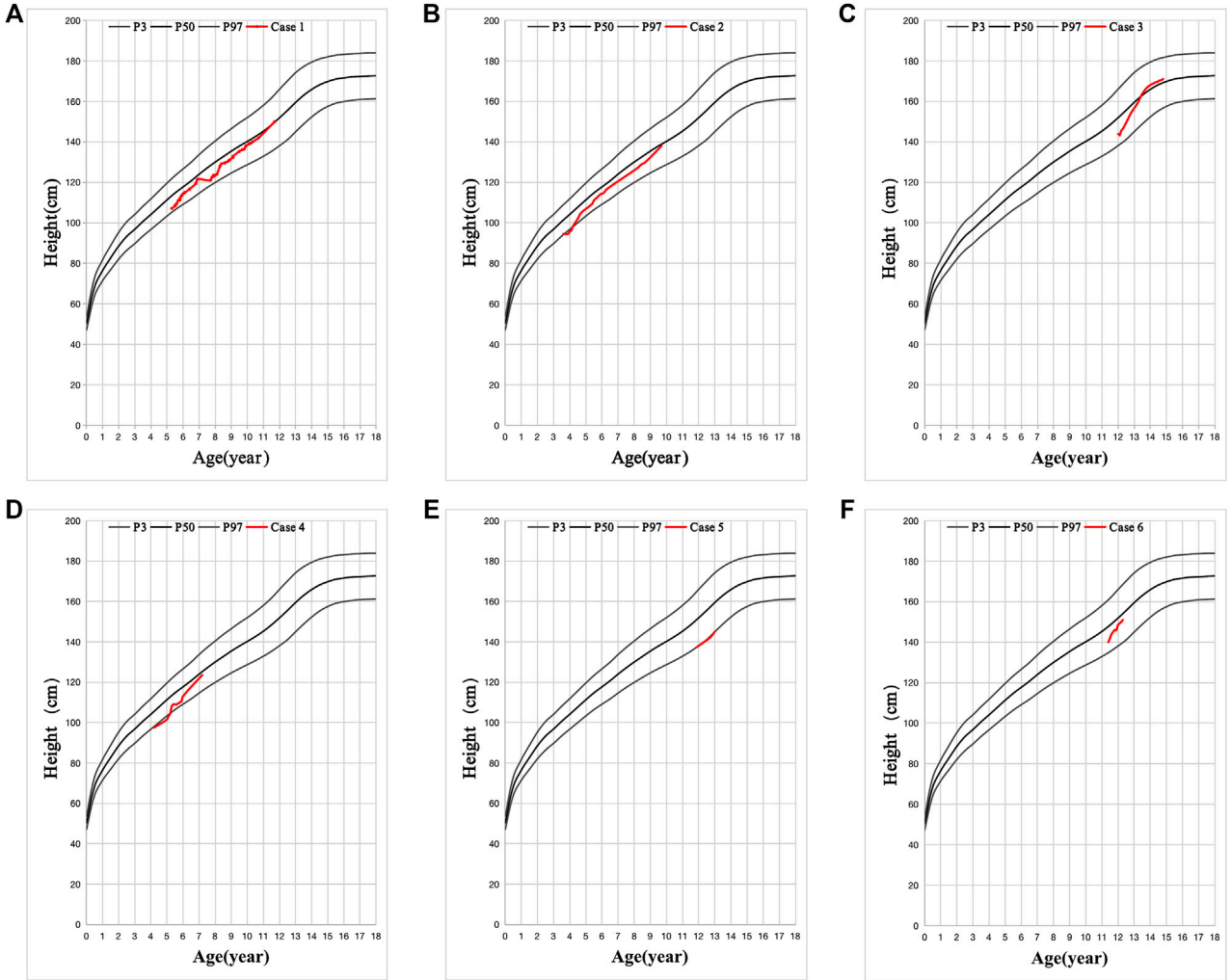


FIGURE 2 | Six cases with favorable effect to recombinant human growth hormone (rhGH) treatment. **(A-F)** is the height changes in six cases. The black line is the standard height curve from top to bottom is 97, 50, and 3%, respectively. Height below the 3rd percentile is considered as Short Stature. The red line is the height changes of these 6 patients after VitD3 and rhGH treatment.

transcription factors (Khan et al., 2018). To our surprise, 11 transcription factors (MEF2C, ZNF384, SOX15, LM140, SOX15, SOX10, RORC, RORA, RORB, NR4A1, NR2F2, and NR4A2) could bind to the regions containing the novel variant under default parameters.

DISCUSSION

Individual children often have multisystem diseases. The clinician needs to be aware of these to ensure that the child is not misdiagnosed. In the past, short stature and obesity in children with allergic diseases are usually considered to be side effects of glucocorticoid drugs. But now, our study generated evidences showing that short stature and central obesity are not related to the use of glucocorticoid drugs. Therefore, we need a new theory to explain the pathological mechanisms of short stature and/or obesity with allergies in children. In this report, the triad of allergic march, short stature, and fatty liver is associated with a patrilineal imprinted gene *SLC22A18*. It should not simply be considered a side effect of glucocorticoid.

There is a progression, and the individual child can suffer from one symptom to another. The most common is normal height and weight at birth. In early childhood, allergies, reduced growth rate, short stature, and abnormal fat metabolism [high FFA and high apolipoprotein E (ApoE)] gradually appear. Around adolescence, weight gain develops into central obesity and fatty liver. And allergic symptoms almost always appear after birth to the age of 7. Other symptoms of itching, sneezing, loss of sleep, coughing, etc., are present. In our study, there is no obvious abnormality in the appearance of our cases, and there is no sexual developmental delay and intellectual disability in any of these patients. The group of six patients could not be classified into the above syndromes.

Several syndromes involving short stature are associated with a number of imprinted genes, such as Prader–Willi syndrome

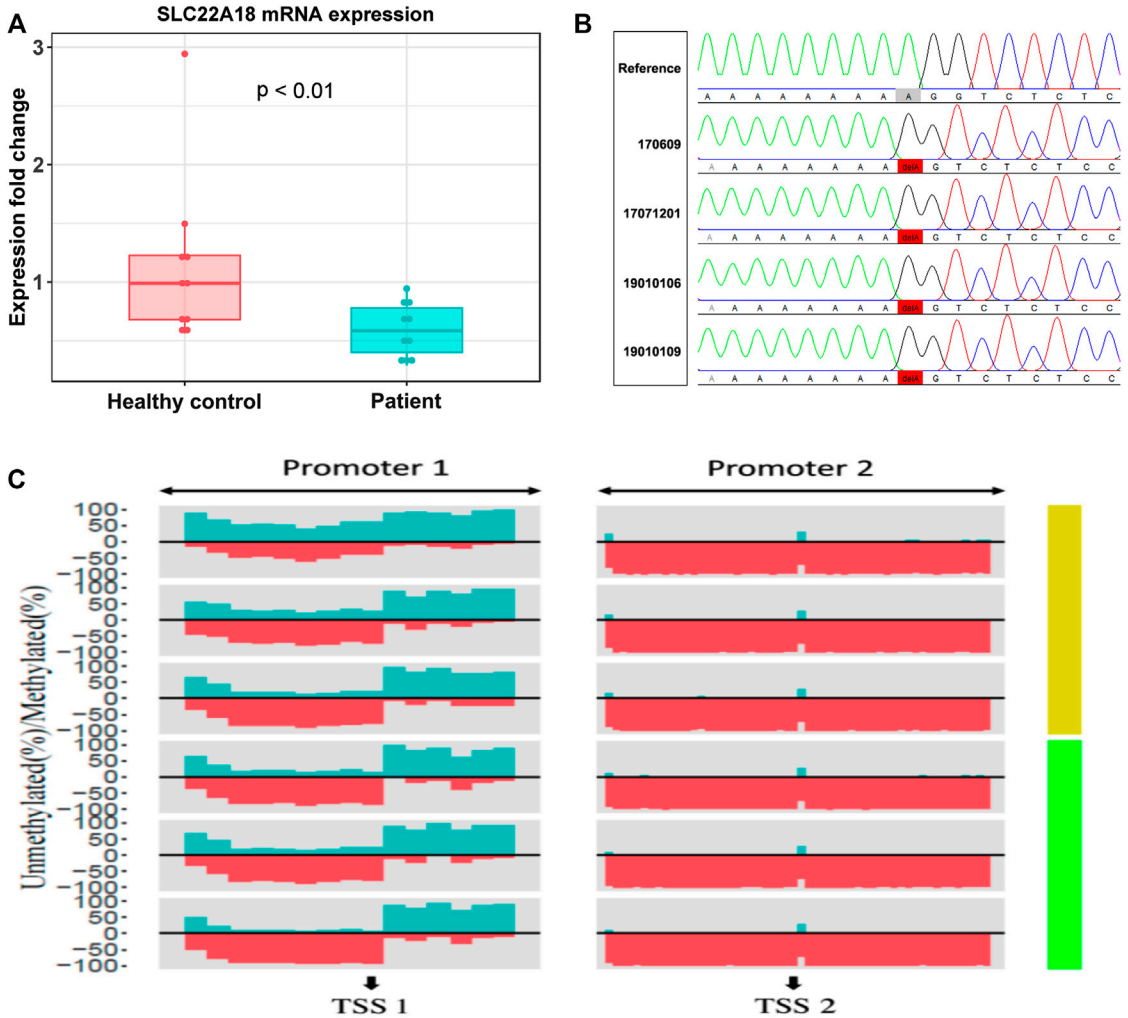


FIGURE 3 | The focal point of a patrilineal imprinted gene *SLC22A18*. **(A)** Comparison of SLC22A18 expression levels between patients and healthy controls. There are 10 individuals in each group, and the expression levels were measured by RT-qPCR, and fold changes were calculated with the healthy controls as the reference. The *p*-value indicated in the plot was based on a two-tailed Student's t-test in the R language. **(B)** The left panel shows the IDs of patients with this variant, and the reference sequence is on the top. The solid red box highlights the variant where an "A" has been deleted. The visualization of Sanger sequencing results was done by novoSNP. **(C)** The percentage of methylated and methylated CpG sites was represented by dark green and red, respectively. Patients and healthy controls were represented by light green and yellow bars, respectively, on the right side. "TSS1" is located at position 2, 899, 721, "+" strand, chromosome 11; "TSS2" is located at position 2, 902, 282, "+" strand, chromosome 11.

(PWS), Beckwith–Wiedemann syndrome (BWS), and Silver–Russell syndrome (SRS). PWS is characterized by short stature and obesity, somewhat similar in our patients. However, inborn muscular hypotonia, imbecility, and cryptorchidism, and/or micropenis are not observed in our cases (Cassidy et al., 2012). BWS features varying degrees of symptoms like overgrowth, macrosomia, macroglossia, hemihypertrophy, and asymmetric facial features (Frédéric et al., 2018). SRS is characterized by severe intrauterine and postnatal growth retardation, feeding difficulties, short stature, triangular face, low ears, and bending of the fifth finger (Spiteri et al., 2017). The group of six patients could not be classified into the above syndromes.

Short stature is the most confusing common feature in the cases of this study. During follow-up, six cases showed a favorable effect of rhGH treatment. All patients grew faster and taller, but it was almost impossible to avoid allergies, consistently elevated plasma levels of IgE, FFA, and fatty liver formation. The molecular mechanism is still unknown, and rare evidence can be found by existing researches.

SLC22A18, located in the 11p15.5 domain, is an important tumor-suppressor gene region. Alterations in this region have been associated with the BWS, Wilms tumor, and lung, ovarian, and breast cancers. Lee et al. found mutations in SLC22A18 in kidney and lung cancers. By checking the genotypes and phenotypes of the family members, the author speculated that SLC22A18 is a tumor suppressor gene in the adult lung and an imprinted tumor suppressor gene in the fetal kidney (Lee et al., 1998). Recent studies revealed a novel link between *SLC22A18* and fat accumulation, and we speculated that paternal *SLC22A18* gene may be involved in the occurrence of this "triad" syndrome. Previous reports found that the liver expresses *SLC22A18* at the mRNA and protein levels (Dao et al., 1998). Suppression of

SLC22A18 promotes lipid accumulation in the liver by reducing lipophagy (Jing and Gotoda, 2012; Yamamoto et al., 2013; Shingo et al., 2019). Dysregulation of lipid metabolism in the liver is a marker of nonalcoholic fatty liver disease (NAFLD), characterized by excessive accumulation of fat in the liver. These findings may explain the development of fatty liver in our cases.

Given these findings and the biological importance of *SLC22A18*, the DNA sequence and the RNA levels of *SLC22A18* were investigated among these patients. It was found that high methylation and low expression of SLC22A18 could relate to the occurrence of a triad of slow growth, allergies, and fatty liver in these patients during their growth and development. Therefore, it is suggested that *SLC22A18* is a possible gene that plays an important role in the pathogenesis of this syndrome.

In summary, this study found that the triad, variable allergy, short stature, and fatty liver, is associated with the lower mRNA expression levels of SLC22A18, deleted "A" in SLC22A18 core promoter, and the high methylation levels in my cases. All of that possibly affect the normal transcription of SLC22A18, meanwhile resulting in IGF-1 low activity and involvement of high FFA in metabolic inflammation. Therefore, our study raises the clinical need for the naming of *SLC22A18* syndrome. Last but not least, additional case samples are needed to reinforce our hypothesis; and further researches of *SLC22A18*, such as epigenetics or functional genetic experiments of an animal model, will reveal the molecular mechanisms and etiology.

DATA AVAILABILITY STATEMENT

The Data Availability statement is corrected to "The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ebi.ac.uk/arrayexpress/, E-MTAB-9579.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the East Hospital Biobank. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JQ contributed to the overall clinical observation, clinical management, follow-up, experimental design, data analysis, and summary of this study. JQ wrote this article. YC performed most of the experiments. YC and XL recorded and collected clinical data, and YL helped with the clinical diagnosis. WQ and AL played a role in patient recruitment. TW played a role in the rhGH treatment. YC and GC have made reasonable suggestions for this study. All authors have read and approved the content and agreed to submit it for consideration for publication in the journal.

FUNDING

This research was supported by the Shanghai Science and Technology Committee (No. 14411971400) and Natural Science Foundation of Jiangxi Province (No. 20171BAB205071).

ACKNOWLEDGMENTS

Biospecimens were obtained from East Hospital Biobank. All patients had signed informed consent for donating their specimens to Shanghai East Hospital Biobank. Approval for using human blood was obtained from the Review Board (IRB) Committee of the institutes mentioned above. We are grateful for the willingness of the patients and their parents to contribute to this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.784135/full#supplementary-material

REFERENCES

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., et al. (2012). methylKit: a Comprehensive R Package for the Analysis of Genome-wide DNA Methylation Profiles. *Genome Biol.* 13, R87. doi:10.1186/gb-2012-13-10-r87

Cassidy, S. B., Schwartz, S., Miller, J. L., and Driscoll, D. J. (2012). Prader-Willi Syndrome. *Genet. Med.* 14, 10–26. doi:10.1038/gim.0b013e31822bead0

Dao, D., Frank, D., Qian, N., O'Keefe, D., Vosatka, R. J., Walsh, C. P., et al. (1998). IMPT1, an Imprinted Gene Similar to Polyspecific Transporter and Multi- Drug Resistance Genes. *Hum. Mol. Genet.* 7, 597–608. doi:10.1093/hmg/7.4.597

FastQC (2010). A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. (Accessed 2010).

Frédéric, B., Kalish, .J. M., Alessandro, M., Foster, A. C., Jet, B., Battista, F. G., et al. (2018). Expert Consensus Document: Clinical and Molecular Diagnosis, Screening and Management of Beckwith-Wiedemann Syndrome: an International Consensus Statement. *Nat. Rev. Endocrinol.* 14, 229–249. doi:10.1530/ey.15.2.2

Ito, S., Honda, G., Fujino, Y., Ogata, S., Hirayama-Kurogi, M., and Ohtsuki, S. (2019). Knockdown of Orphan Transporter SLC22A18 Impairs Lipid Metabolism and Increases Invasiveness of HepG2 Cells. *Pharm. Res.* 36, 39. doi:10.1007/s11095-018-2565-4

Jing, I. Y. Q., and Gotoda, K. (2012). Elucidate to the Mechanism of SLC22A18 Gene Associated with the Adioposis Accumulation. *Ther. Res.* 33, 838–842.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., et al. (2018). JASPAR 2018: Update of the Open-Access Database of Transcription Factor Binding Profiles and its Web Framework. *Nucleic Acids Res.* 46, D260–d266. doi:10.1093/nar/gkx1126

Krueger, F., and Andrews, S. R. (2011). Bismark: a Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications. *Bioinformatics* 27, 1571–1572. doi:10.1093/bioinformatics/btr167

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., et al. (2014). ClinVar: Public Archive of Relationships Among Sequence Variation and Human Phenotype. *Nucl. Acids Res.* 42, D980–D985. doi:10.1093/nar/gkt1113

Lee, M. P., Reeves, C., Schmitt, A., Su, K., Connors, T. D., Hu, R. J., et al. (1998). Somatic Mutation of TSSC5, a Novel Imprinted Gene from Human Chromosome 11p15.5. *Cancer Res.* 58, 4155–4159.

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *Embnet J.* 17, 10. doi:10.14806/ej.17.1.200

Palomares, O., Akdis, M., Martín-Fontecha, M., and Akdis, C. A. (2017). Mechanisms of Immune Regulation in Allergic Diseases: the Role of Regulatory T and B Cells. *Immunol. Rev.* 278, 219–236. doi:10.1111/imr.12555

Peters, J. (2014). The Role of Genomic Imprinting in Biology and Disease: an Expanding View. *Nat. Rev. Genet.* 15, 517–530. doi:10.1038/nrg3766

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI Database of Genetic Variation. *Nucleic Acids Res.* 29, 308–311. doi:10.1093/nar/29.1.308

Shingo, I., Gentaro, H., Yu, F., Seiryo, O., Mio, H.-K., and Sumio, O. (2019). Knockdown of Orphan Transporter SLC22A18 Impairs Lipid Metabolism and Increases Invasiveness of HepG2 Cells. *Pharm. Res.* 36, 39. doi:10.1007/s11095-018-2565-4

Sohi, D. K., and Warner, J. O. (2008). Understanding Allergy. *Paediatrics Child. Health* 18, 301–308. doi:10.1016/j.paed.2008.04.006

Spiteri, B. S., Stafrace, Y., and Calleja-Agius, J. (2017). Silver-Russell Syndrome: A Review. *Neonatal. Netw.* 36, 206–212. doi:10.1891/0730-0832.36.4.206

Weckx, S., Del-Favero, J., Rademakers, R., Claes, L., Cruts, M., De Jonghe, P., et al. (2005). novoSNP, a Novel Computational Tool for Sequence Variation Discovery. *Genome Res.* 15, 436–442. doi:10.1101/gr.2754005

Yamamoto, T., Izumi-Yamamoto, K., Iizuka, Y., Shirota, M., Nagase, M., Fujita, T., et al. (2013). A Novel Link between Slc22a18 and Fat Accumulation Revealed by a Mutation in the Spontaneously Hypertensive Rat. *Biochem. Biophys. Res. Commun.* 440, 521–526. doi:10.1016/j.bbrc.2013.09.096

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 February 2022
doi: 10.3389/fgene.2021.798001

Association Between a *TLR2* Gene Polymorphism (rs3804099) and Proteinuria in Kidney Transplantation Recipients

Shuang Fei[1†], Zeping Gui[1,2†], Dengyuan Feng[1], Zijie Wang[1], Ming Zheng[1], Hao Chen[1], Li Sun[1], Jun Tao[1], Zhijian Han[1], Xiaobing Ju[1], Min Gu[2], Ruoyun Tan[1]* and Xinli Li[3]**

[1]Department of Urology, The First Affiliated Hospital with Nanjing Medical University, Nanjing, China, [2]Department of Urology, The Second Affiliated Hospital with Nanjing Medical University, Nanjing, China, [3]Department of Cardiology, The First Affiliated Hospital with Nanjing Medical University, Nanjing, China

OPEN ACCESS

Edited by:
Yuriy L. Orlov,
I.M. Sechenov First Moscow State Medical University, Russia

Reviewed by:
Hae Jeong Park,
Kyung Hee University, South Korea
Xiuqin Yang,
Northeast Agricultural University, China

***Correspondence:**
Min Gu
lancetgu@aliyun.com
Ruoyun Tan
tanruoyun112@vip.sina.com
Xinli Li
xinli3267@njmu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 19 October 2021
Accepted: 31 December 2021
Published: 21 February 2022

Citation:
Fei S, Gui Z, Feng D, Wang Z, Zheng M, Chen H, Sun L, Tao J, Han Z, Ju X, Gu M, Tan R and Li X (2022) Association Between a TLR2 Gene Polymorphism (rs3804099) and Proteinuria in Kidney Transplantation Recipients. Front. Genet. 12:798001. doi: 10.3389/fgene.2021.798001

Background: The occurrence of proteinuria is one of the evaluation indicators of transplanted kidney damage and becomes an independent risk factor for poor prognosis after kidney transplantation. Our research sought to understand these potential associations and detect the underlying impact of single-nucleotide polymorphisms (SNPs) on proteinuria in kidney transplant recipients.

Materials and Methods: There were 200 recipients enrolled in this study, from which blood samples were extracted for SNP mutation–related gene detection. RNA sequencing was performed in kidney tissues after kidney transplantation, and the significantly differentially expressed genes (DEGs) were analyzed between the control group and the proteinuria group. Then, the intersection of genes with SNP mutations and DEGs was conducted to obtain the target genes. Multiple genetic models were used to investigate the relationship between SNPs and proteinuria. In addition, the effect of SNP mutation in the target gene was further validated in human renal podocytes.

Results: According to the sequencing results, 26 significant SNP mutated genes and 532 DEGs were found associated with proteinuria after kidney transplantation. The intersection of SNP mutated genes and DEGs showed that the Toll-like receptor 2 (TLR2) gene was significantly increased in the transplanted renal tissues of patients with proteinuria after kidney transplantation, which was consistent with the results of immunohistochemical staining. Further inheritance model results confirmed that mutations at rs3804099 of the TLR2 gene had significant influence on the occurrence of proteinuria after kidney transplantation. In the *in vitro* validation, we found that, after the mutation of rs3804099 on the TLR2 gene, the protein expressions of podocalyxin and nephrin in podocytes were significantly decreased, while the protein expressions of desmin and apoptosis markers were significantly increased. The results of flow cytometry also showed that the mutation of rs3804099 on the TLR2 gene significantly increased the apoptotic rate of podocytes.

Conclusion: Our study suggested that the mutation of rs3804099 on the TLR2 gene was significantly related to the generation of proteinuria after kidney transplantation. Our data provide insights into the prediction of proteinuria and may imply potential individualized therapy for patients after kidney transplantation.

Keywords: kidney transplantation, proteinuria, single-nucleotide polymorphisms, high-throughput sequencing, Toll-like receptor 2

BACKGROUND

With the improvement of organ transplantation methods and the application of new immunosuppressive agents, the short-term survival rate of transplanted kidney has been significantly ameliorated. However, the long-term survival rate of transplanted kidney needs to be improved, and the long-term complications after kidney transplantation are receiving more attention. Consistent with proteinuria in common chronic kidney diseases, persistent proteinuria after kidney transplantation is closely connected with the prognosis of kidney transplantation (Blasius and Beutler, 2010). The incidence of proteinuria 1 year after kidney transplantation is 15–20%. The long-term survival of transplanted kidney is directly related to the occurrence of proteinuria, which is an independent risk factor affecting allograft survival and leads to the death of patients after transplantation (Fernandez-Fresnedo et al., 2004). Immune and non-immune factors after kidney transplantation contribute to proteinuria (Sancho et al., 2007). Studies have shown that enhanced immunosuppressive therapy is ineffective for post-transplantation proteinuria. The generation of proteinuria after kidney transplantation attributes to chronic rejection, chronic graft nephropathy, glomerulonephritis after kidney transplantation, acute rejection, and cyclosporine nephrotoxicity (Reichel et al., 2004). Effective control of proteinuria after transplantation can reduce the damage to the transplanted kidney.

Nowadays, abundant research studies have demonstrated that gene polymorphisms are associated with kidney disease. The mutation and polymorphism of nephrin gene NPHS1 impact the development of congenital nephrotic syndrome of Finnish type (CNF), MCNS, and other diseases as well as the occurrence of its proteinuria. Mutations in the NPHS2 gene of podocin contribute to steroid-resistant nephrotic syndrome (SRNS), most of which are familial but may also are sporadic (Rood et al., 2019). Mutations in the podocin gene are also relevant to non-diabetic end-stage renal disease in African Americans (Menara et al., 2020). And clear correlation between the polymorphism of this gene and the occurrence of proteinuria has been validated in the general population. This evidence all indicates that genetic mutations obviously have an influence on the pathogenesis of kidney disease.

Toll-like receptors (TLRs), a family of structural recognition receptors across cell membranes, play an important role in inflammation, immune response, and tumorigenesis, where TLR2 is a crucial member. The coding gene of TLR2 is located on the 4q32 region of chromosome 4, and TLR2 is mainly expressed on the surface of peripheral blood leukocytes. TLR2 can interact with bacteria and endogenous ligands to activate transcription factors, such as NF-κB and AP-1 (Harding and Boom, 2010; Vilahur and Badimon, 2014). Current studies have validated that, after tissue ischemic injury, numerous apoptotic cells could emerge in the organs and generate a series of immune responses such as host defense and injury repair responses (Marshak-Rothstein, 2006; Obhrai and Goldstein, 2006; Seki and Brenner, 2008). The accumulation of inflammatory cells along with the release of inflammatory factors can damage the tissues. The loss of TLR2 can lead to kidney damage, leukocyte influx, and renal tubular damage during renal ischemia–reperfusion. Ding et al. (2015) reported that proteinuria might exhibit an endogenous danger-associated molecular pattern (DAMP) that induced tubulointerstitial inflammation *via* TLR2–MyD88–NF-κB pathway activation. In this study, the correlation between TLR2-regulated gene expressions and proteinuria after kidney transplantation will be investigated.

MATERIALS AND METHODS

Ethics Statement

All research design, patient registration, and procedure agreement were approved by the local ethics committee of the First Affiliated Hospital of Nanjing Medical University (2016-SR-029). All kidney-transplanted recipients expressed their knowledge and understanding of the experimental process and provided written informed consent at the same time. All our research processes follow the ethical standards of the "Helsinki Declaration" and the "Istanbul Declaration."

Study Design and Population

We use a single-center, retrospective, cohort study method to explore the impact of single-nucleotide polymorphisms of genes based on TLR2-related signaling pathways and the progress of proteinuria in kidney-transplanted recipients. A total of 200 recipients who received kidney transplanted in the Kidney Transplant Center at the First Affiliated Hospital of Nanjing Medical University from February 1, 2015, to September 1, 2018, participated in this study. There is no significant transplanted renal function failure or decline during the current follow-up. The detailed research methods including the inclusion and exclusion criteria were described in detail in our previous studies (Wang et al., 2018). The clinical data, including age, gender, height, independence, and immunosuppressive protocol, were recorded by one of the authors, Zeping Gui.

Immunosuppressive Protocols

Our center selected triple immunosuppressive regimen [cyclosporin A or tacrolimus, combined with mycophenolate mofetil (MMF) and prednisone] or quadruple immunosuppressive regimen (cyclosporin A or tacrolimus, prednisone, MMF combined with sirolimus) as maintenance-period immunosuppressant treatment. We adjusted the dosage of the corresponding immunosuppressive regimen according to drug concentration levels and blood creatinine level. Our previous research studies had reported corresponding details about detailed information and methodologies for immunosuppressive agent schedules (Wang et al., 2018).

Sample Collection and Next-Generation Sequencing (NGS)

Peripheral blood (2 ml) of each patient was used for DNA extraction. We quantitatively analyzed the concentration and purity of genomic DNA (gDNA) and assessed the integrity of genes by agarose gel electrophoresis. We chose target-specific target regions from a random pool containing upstream and downstream oligonucleotides and gDNA hybrids. Then, we dispersed gDNA and selectively amplified adapter-ligated DNA, limited-cycle polymerase chain reaction (PCR). We denatured the captured library and loaded it into an Illumina cBot instrument manufacturer's protocol. Subsequently, we analyzed the sequencing data based on the human available data reference sequence UCSC hg19 assembly (NCBI construction 37.2), using the genome analysis tool, Picard software, dbSNP 132. We also found two separate programs for recognizing somatic mutation cells: MuTect 1.1.5 and VarScan 2.3.6.

Kidney Transplant Tissue Samples and RNA Sequencing Methods

We collected kidney transplant tissues in three recipients who were subjected to transplanted kidney biopsy due to proteinuria operated at First Affiliated Hospital with Nanjing Medical University between 2016 and 2018. In addition, three normal kidney samples were obtained from patients undergoing radical nephrectomy, and each sample was excised at 5 cm away from tumor tissue. The collected samples were stored at −80°C for the preparation of RNA extraction. And all the samples were analyzed separately.

Total RNA was used as an input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, United States) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo–attached magnetic beads. Fragmentation was carried out using divalent cations at elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First strand cDNA was synthesized using the random hexamer primer and M-MuLV Reverse Transcriptase (RNase H). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends *via* exonuclease/polymerase activities. After adenylation of 3′ ends of DNA fragments, NEBNext Adaptor with a hairpin loop structure was ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 250–300 bp in length, the library fragments were purified with an AMPure XP system (Beckman Coulter, Beverly, United States). Then, 3 μl USER Enzyme (NEB, United States was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. Then, PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. At last, PCR products were purified (AMPure XP system), and library quality was assessed on the Agilent Bioanalyzer 2100 system.

Identification and Elucidation of Differentially Expressed Genes (DEGs)

DEGs were screened with |log fold-change (FC)| >1 and p-value < 0.05 between the proteinuria and normal groups, using the limma R package (version 3.13). The Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the identified DEGs were performed on DAVID 6.8 (https://david.ncifcrf.gov/). p-Value<0.05 was considered statistically significant.

Immunohistochemistry (IHC) Staining Assay

Kidney samples were fixed in 10% neutral formalin and embedded in paraffin. Paraffin sections (3 μm) were deparaffinized, hydrated, and antigen-retrieved. The endogenous peroxidase activity was quenched by 3% H_2O_2. Sections were then blocked with 10% normal donkey serum, followed by incubation with anti-TLR2 (1:100; Abcam, United States) overnight at 4°C. After incubation with biotinylated goat anti-mouse/rabbit IgG (0.5 μg/ml; Abcam) for 1 h, sections were incubated with ABC reagents for 1 h at room temperature before being subjected to substrate 3-amino-9-ethylcarbazole or 3,3′-diaminobenzidine (Vector Laboratories, Burlingame, CA). Slides were viewed under a Nikon Eclipse 80i microscope equipped with a digital camera (DS-Ri1, Nikon, Shanghai, China).

Cell Culture and Transfection

Human podocytes cryopreserved at −80°C were resuscitated with 10% FBS-containing RPMI 1640 medium, maintained at 33°C (permissive conditions) and 5% CO2, and cultured at 37°C (non-permissive conditions). After 14 days under non-permissive conditions, the cells revealed an arborized shape. Podocytes digested with 0.25% trypsin for passage after differentiating and passaged according to cell growth once every 2–3 days.

Before transfection, human podocytes were starved with FBS-free RPMI 1640 medium overnight. Then, cell transfection was performed according to the manufacturer's protocol.

Plasmid purification, restriction digestion, endonuclease digestions, gel electrophoresis, PCR, ligation, and *E. coli* transformations were carried out by Hanheng Biotechnology

(Hanheng Biotechnology Co., Ltd., Shanghai, China) (**Supplementary Table S1**). Briefly, we designed the primer sequences of rs3804099 mutation and TLR2 wild-type, respectively. The digested products were electrophoresed to separate restriction fragments in agarose gel. The digested vector and PCR fragment were ligated and used to transform DH5α competent cells. The two target fragments (wild-type and SNP mutant type) were ligated to the vector pcDNA3. Products were transformed and propagated in DH5α competent cells. Then, they were grown in lysogeny broth (LB) or on LB agar plates and incubated in an incubator at 37°C for 12 h. The monoclonal colonies were selected for amplification and identified by bacterial liquid PCR. Final plasmid constructs were confirmed by DNA sequencing. The concentration of the plasmids was more than 200 ng/μl. The final derived plasmids are referred to as pCDNA3.1 and pCMV-TLR2 (mut) and used for subsequent cell experiments. In our previous study, we have validated that transfection of an empty plasmid is identical to that of the wild-type plasmid (Wang et al., 2020).

Flow Cytometry Detecting FITC-Annexin V Positive Apoptotic Cells

The cell apoptosis was detected by the FITC-Annexin V Apoptosis Detection Kit (Cat#556547, BD Pharmingen) as described previously. Briefly, the cells with indicated treatment were stained with FITC-Annexin V and propidium iodide (PI). Both early (Annexin V+/PI-) and late (Annexin V+/PI+) apoptotic cells were sorted by fluorescence-activated cell sorting (FACS) (Beckman Coulter Inc., Brea, CA).

Western Blot Analysis

To extract the total proteins, cells were washed twice with cold PBS and lysed in lysis buffer (0.2% SDS, 1% NP-40, 5 mM EDTA, 1mM PMSF, 10 g/ml leupeptin, and 10 g/ml aprotinin) after treatment. Lysates were centrifuged at 4°C for 20 min at 1,500 g. Protein concentrations were determined by a Bradford assay (Bradford, 1976). Afterward, the proteins were separated by 12% SDS-polyacrylamide gel electrophoresis (PAGE) and transferred onto polyvinylidene difluoride (PVDF) membranes. Membranes were blocked with 5% non-fat milk in TBS (pH 7.4) with 0.1% Tween-20 for 1 h at room temperature and then incubated overnight at 4°C with different primary antibodies. The signals were detected by HRP-conjugated secondary antibodies for 1 h at room temperature on an ECL detection system (Amersham Biosciences). To isolate proteins in the cytoplasm and nucleus, the Nucleus Protein Extraction Kit was used according to the instructions (Boster, Wuhan, China). The primary antibodies were listed as follows: anti-TLR2 (1:1,000; Abcam, United States), anti-Desmin (1:1,000; Abcam, United States), anti-Podocalyxin (1:1,000; Abcam, United States), anti-Nephrin (1:1,000; Abcam, United States), anti-GAPDH (1:1,000; CST, United States), anti-Bcl-2 (1:1,000; CST, United States), anti-Bax (1:1,000; CST, United States), and anti-Cleaved Caspase3 (1:1,000; CST, United States). Proteins in the cytoplasm and nucleus were checked by western blot assays and were normalized to GAPDH. The relative intensities of the signals were quantified by densitometry and imaging software (ImageJ, National Institutes of Health, United States).

TABLE 1 | Basic demographics of kidney transplantation patients in this cohort.

Characteristics	PKT (*n*)	Non-PKT (*n*)
Case number	97	103
Weight (kg; mean ± SD)	60.09 ± 9.21	60.94 ± 9.03
Recipient age (years; mean ± SD)	37.31 ± 9.98	35.65 ± 8.92
Gender (male/female)	72/25	69/34
PRA before renal transplant (%)	0	0
Primary/secondary renal transplant		
Primary renal transplant	97	103
Secondary renal transplant	0	0
Type of donor		
Living donor	15	9
DCD	82	94
Administration of sirolimus (%)	11.34	12.62
DGF (%)	37.11	29.13

Abbreviation: SD, standard deviation; PRA, panel reactive antibody; DCD, donor after cardiac death; DGF, delayed graft function; PKT, proteinuria after kidney transplantation.

Statistical Analysis

Unless otherwise stated, data were expressed as mean ± standard deviation (SD). The minor allele frequency (MAF) and Hardy–Weinberg equilibrium (HWE) were conducted by R package genetics (genetics: Population Genetics, R package version 1.3.8.1.). The linkage disequilibrium (LD) blocks were analyzed using Haploview version 4.2 (Broad Institute, Cambridge, MA, United States). General linear models (GLMs) were used to determine the importance and influence of clinical variables on proteinuria. We used the R Statistics Package SNPassoc (SNP-based whole genome association studies; R package version 1.9-2.) to examine five multiple inheritance models [codominant model 1 (major allele homozygotes versus heterozygotes), codominant model 2 (major allele homozygotes versus minor allele homozygotes), dominant model (major allele homozygotes versus minor allele homozygotes plus heterozygotes), recessive model (major allele homozygotes plus heterozygotes versus minor allele homozygotes), log-additive model (major allele homozygotes versus heterozygotes versus minor allele homozygotes), and overdominant model (heterozygotes versus major allele homozygotes plus minor allele homozygotes)]. All data in our study were analyzed using SPSS software version 13.0 (SPSS Inc., Chicago, IL, United States), and $p < 0.05$ was considered statistically significant.

RESULTS

Patient Demographics

A total of 97 patients with proteinuria and 103 control kidney-transplanted patients were recruited in this study. The clinical characteristics of the patients, including weight, recipient age, gender, incidence of PRA before kidney transplant, primary or

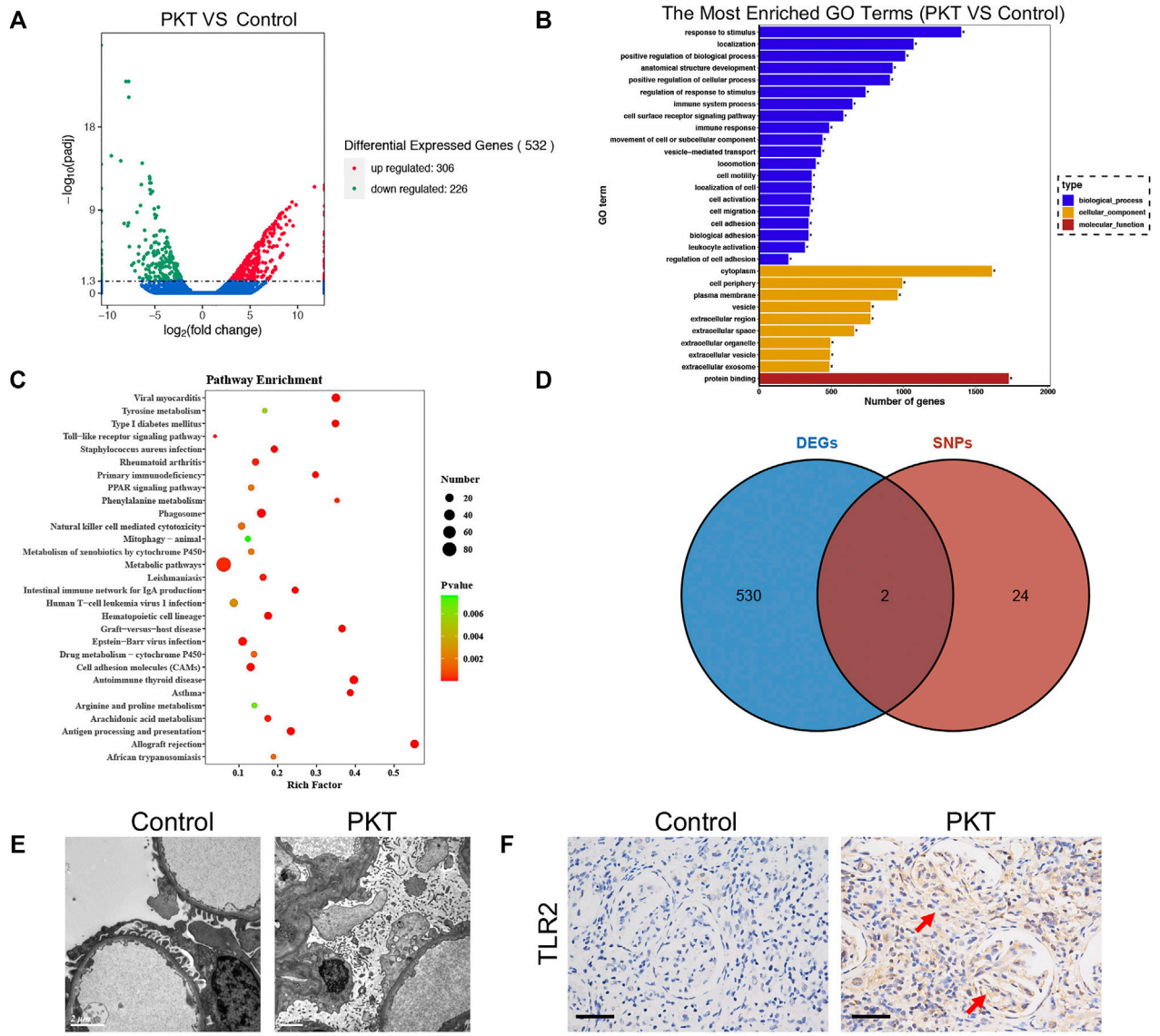


FIGURE 1 | Identification and functional annotation of DEGs and validation of TLR2 expression in tissue. **(A)** Volcano plots of 532 DEGs are shown between the PKT and control groups with 306 upregulated genes and 226 downregulated genes. **(B)** GO enrichment analysis–depicted DEGs are associated with the immune system and inflammatory response. **(C)** KEGG analysis of the DEGs showing the top 20 enriched KEGG terms including toll–like receptor signaling pathway. **(D)** Venn diagram exhibiting two intersected genes including the TLR2 gene between DEGs and SNP mutation–related genes. **(E)** Podocyte foot effacement in podocytes of glomeruli is seen in the PKT group under electron microscopy compared with the control group. **(F)** IHC staining showing TLR2 expression is higher in the PKT group than the control group. Abbreviation: DEGs, differentially expressed genes; PKT: proteinuria after kidney transplantation; GO, Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; SNP, single-nucleotide polymorphism; IHC: immunohistochemistry.

secondary kidney transplant, type of donor, administration of sirolimus, and the presence of delayed graft function (DGF), were collected and compared, as illustrated in **Table 1**. The study population age ranged from 28 to 52 years, and the mean age of the patients was 36.45 ± 9.46 years. Of the patients with kidney transplant, 70% were male, while 30% were female. None of the patients presented with PRA before kidney transplant. The patients with living donor formed 12%, and the patients with donor after cardiac death (DCD) formed 88%. Twelve patients were treated with sirolimus.

Tagger SNP Selection

We identified SNPs of 26 genes (**Supplementary Table S2**) for further analysis using HWE analysis and filter criteria of MAF>0.05. These SNPs were deemed as statistically frequent, whereas the remaining were identified to be rare. As shown in **Figure 1A**, DEGs between control and PKT (proteinuria after kidney transplantation) groups (n = 3) were evaluated by RNA sequencing. Red represents an upregulated gene, and green represents a downregulated gene. By assessing the sequencing results, 532 genes including 306 upregulated and 226

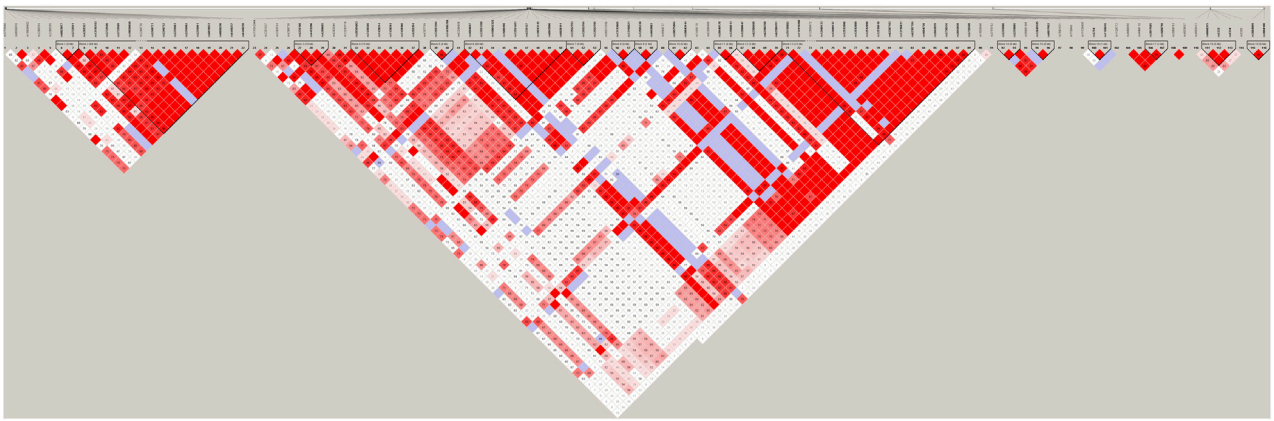


FIGURE 2 | LD results of detected SNPs in the TLR2 gene. Abbreviation: LD, linkage disequilibrium; SNPs, single-nucleotide polymorphisms.

downregulated genes were found differentially expressed between the two groups, based on the screening condition of $|\log_2$ fold-change$|>1$ and p-value < 0.05. GO enrichment analysis suggested that these DEGs were related to the immune system and inflammatory response, such as cell adhesion (GO:0007155) and regulation of cell adhesion (GO:0030155) (**Figure 1B**). KEGG pathway enrichment analysis of the DEGs showed the top 20 enriched KEGG terms, including toll–like receptor signaling pathway, natural killer cell mediated cytotoxicity, and mitophagy (**Figure 1C**). In addition, two genes including TLR2 and SLC2A9 were selected from the intersection between the 532 DEGs and 26 SNP mutated genes (**Figure 1D**). According to the KEGG pathway analysis (**Figure 1C**), toll–like receptor signaling pathway was among the top 20 enriched KEGG terms. Besides, previous studies support that TLR2 is closely associated with proteinuria (Motojima et al., 2010; Ururahy et al., 2012; Bergallo et al., 2017). Considering that, we selected TLR2 for further analysis. Furthermore, LD analysis was performed among the SNPs to discover the tagger SNPs (**Figure 2**). We take the intersection between SNPs selected by LD analysis and frequent SNPs by HWE/MAF analysis. Hence, a couple of tagger SNPs (rs3804099, rs3804100) from the TLR2 gene were used as representative loci for the subsequent statistical analysis.

For further verifying the impact of TLR2 on the development of proteinuria, experiments using human specimens with (PKT group) or without (control group) proteinuria after kidney transplantation were implemented. IHC staining results showed that the expression of TLR2 significantly increased in the PKT group compared with the control group (**Figure 1F**). Additionally, more podocyte foot effacement in podocytes of glomeruli was observed on electron microscopy in the PKT group than the control group (**Figure 1E**).

Confounding Factor Analysis and Multiple Inheritance Model Analysis

Based on various degrees of correlation with proteinuria, confounding factors, including BKV infection, incidence of acute rejection gender, age, weight, ISD protocol, duration after renal transplant, use of rapamycin, and DGF occurrence, were evaluated by GLM analysis. According to the GLM results, duration after kidney transplantation impacted the generation of proteinuria ($p = 0.042$), whereas other factors made no difference to that ($p > 0.05$) (**Table 2**).

TABLE 2 | Influence of confounding factors on the outcomes of proteinuria by a general linear model in this cohort.

Confounding factors	*F* value	*p-Value*
BKV infection	0.829	0.407
AR	0.36	0.719
Gender	1.311	0.190
Age	1.741	0.082
Weight	−1.212	0.225
ISD protocol	−1.134	0.257
Duration after kidney transplantation	2.033	**0.042**
RAPA	−0.055	0.956
DGF	0.563	0.573

Abbreviation: ISD, immunosuppressive drug; RAPA, rapamycin; DGF, delayed graft function. The bold values are statistically significant.

TABLE 3 | Results of multiple inheritance models in rs3804099 adjusted by the administration of sirolimus in five models.

rs3804099	OR (95% CI)	*p*-Value
Codominant model	1.91 (1.01, 3.62)	**<0.001**
Dominant model	2.38 (1.29, 4.41)	**0.005**
Recessive model	7.19 (1.92, 26.97)	**<0.001**
Overdominant model	1.3 (0.72, 2.35)	0.380
Log-additive model	2.48 (1.5, 4.09)	**<0.001**

Abbreviation: OR, odds ratio; CI, confidence interval. The bold values are statistically significant.

A Bonferroni multiple-testing correction was used to adjust the different duration after kidney transplantation of kidney transplant patients (adjusted p-value = .005), and multiple inheritance model analyses were performed to narrow down the candidate SNPs. After massive univariate testing, SNP rs3804099 was found significantly associated with proteinuria after kidney transplantation [codominant model: OR (95% CI) = 1.91 (1.01, 3.62), $p < 0.001$; dominant model: OR (95% CI) = 2.38

TABLE 4 | Distributions and analysis of rs3804099 in patients with proteinuria.

Genotype	Proteinuria degree					*p*-Value
	-	1+	2+	3+	4+	
TC	46	35	13	2	2	**0.005**
TT	54	24	4	1	1	
CC	3	13	2	0	0	

The bold values are statistically significant.

(1.29, 4.41), $p = 0.005$; recessive model: OR (95% CI) = 7.19 (1.92, 26.97), $p < 0.001$; overdominant model: OR (95% CI) = 1.3 (0.72, 2.35), $p = 0.380$; log-additive model: OR (95% CI) = 2.48 (1.5, 4.09), $p < 0.001$] (**Table 3**). This result suggested that the risk of proteinuria was strongly correlated with the rs3804099, rs3804100 locus (**Supplementary Table S3**), compared with other non-significant SNPs ($p > .05$). After searching the literature studies carefully, we found rs3804099 SNP has been

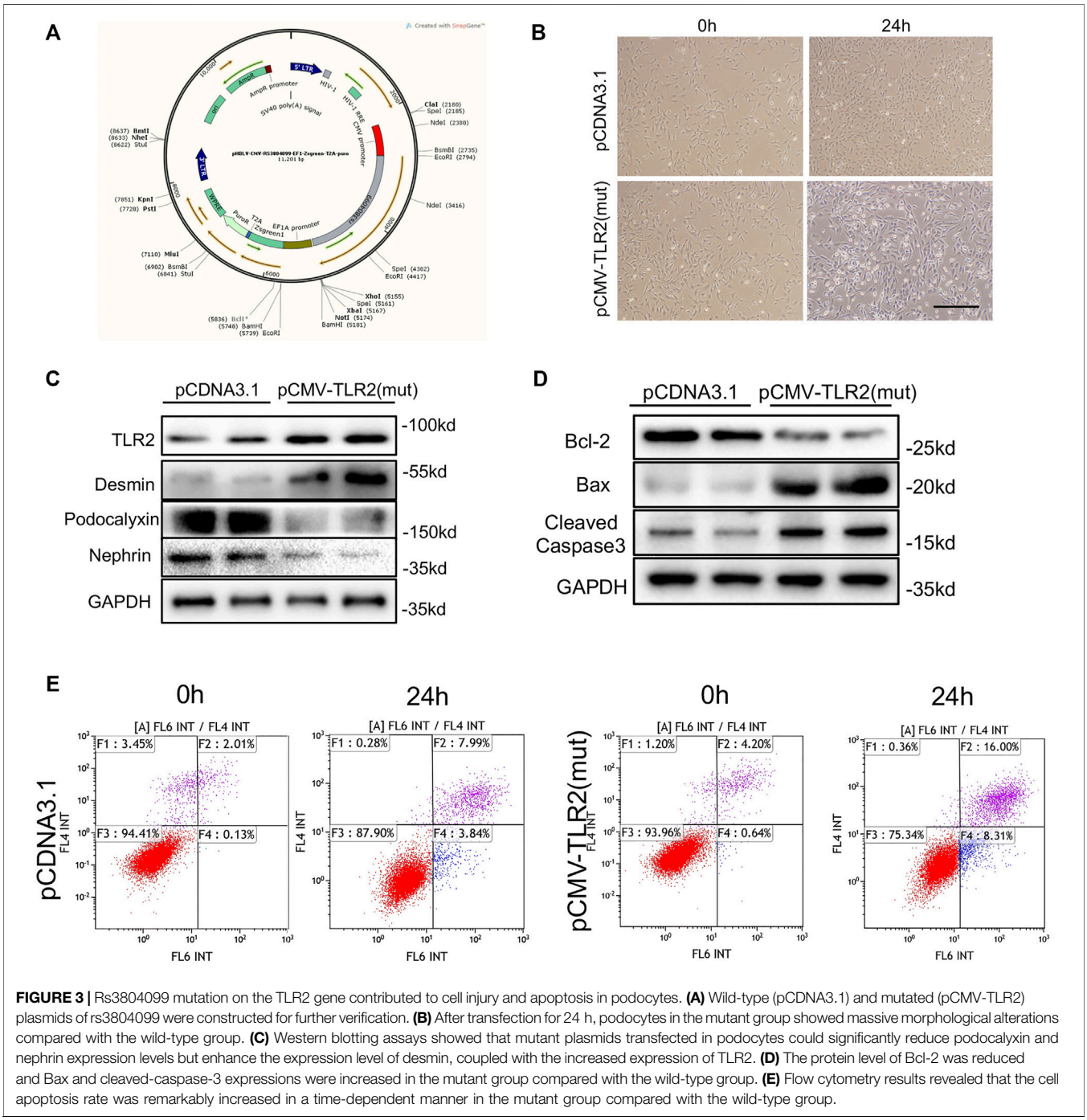


FIGURE 3 | Rs3804099 mutation on the TLR2 gene contributed to cell injury and apoptosis in podocytes. **(A)** Wild-type (pCDNA3.1) and mutated (pCMV-TLR2) plasmids of rs3804099 were constructed for further verification. **(B)** After transfection for 24 h, podocytes in the mutant group showed massive morphological alterations compared with the wild-type group. **(C)** Western blotting assays showed that mutant plasmids transfected in podocytes could significantly reduce podocalyxin and nephrin expression levels but enhance the expression level of desmin, coupled with the increased expression of TLR2. **(D)** The protein level of Bcl-2 was reduced and Bax and cleaved-caspase-3 expressions were increased in the mutant group compared with the wild-type group. **(E)** Flow cytometry results revealed that the cell apoptosis rate was remarkably increased in a time-dependent manner in the mutant group compared with the wild-type group.

reported closely associated with some diseases, including polycystic ovary syndrome (Kuliczkowska-Płaksej et al., 2021) and *Helicobacter pylori* infection and peptic ulcer (Mirkamandar et al., 2018). And rs3804099 polymorphism also has an impact on the susceptibility of cancers (Gao et al., 2019). According to that, we selected rs3804099 for further analysis.

Furthermore, we selected three genotypes and compared differences between the degrees of proteinuria for each of the three groups. The results indicated significant differences (p = 0.005) for degrees of proteinuria among the CC, TC, and TT genotypes (**Table 4**). The analysis results further showed that patients with two T alleles (TT genotypes) showed a significant decreased risk in the development of proteinuria after transplantation.

Rs3804099 Enhanced Expression of TLR2 and Induced Podocyte Injury and Apoptosis

To identify the influence of rs3804099 on podocytes, the pCDNA3 plasmids and the pCDNA3.1 plasmids with rs3804099 mutation were constructed (**Figure 3A**). Then, podocytes transfected with the pCDNA3.1 or pCMV-TLR2 plasmids were assigned to the control or mutant groups, respectively. After transfection for 24 h, podocytes in the mutant group showed massive morphological alterations detected by phase-contrast microscopy (**Figure 3B**). Using the western blotting assays, we assessed the impact of the rs3804099 site and found the increase of TLR2 expression in the rs3804099 mutant group (**Figure 3C**).

Nephrin and podocalyxin are the most important components of the slit membrane of podocytes. Western blotting assays showed that the mutant plasmid transfected in podocytes could significantly reduce podocalyxin and nephrin (podocyte marker proteins) levels but enhance the expression of desmin (a marker of podocyte injury) level, coupled with the increased expression of TLR2. Moreover, the protein level of Bcl-2 was reduced and Bax and cleaved-caspase-3 expressions were increased in the mutant group compared with the wild-type group (**Figure 3D**). Consistently, flow cytometry results revealed that the cell apoptosis rate was remarkably increased in a time-dependent manner in the mutant group compared with the wild-type group (**Figure 3E**).

DISCUSSION

The occurrence of proteinuria is one of the evaluation indicators of kidney transplantation damage, and it is accompanied by renal function damage and becomes an independent risk factor for poor prognosis after kidney transplantation. It has been reported that the quantification and duration of proteinuria after kidney transplantation are positively correlated with the risk of renal allograft function damage and even loss of renal allograft. When proteinuria increases by 100 mg, the risk of renal allograft loss greatly increases (Vachek et al., 2020). The current clinical prevention and treatment measures are reasonable dietary adjustment, blood lipid control, hormones, and immunosuppressive agents to reduce the progressive damage of the kidney's innate cells, but the effect is not satisfactory (Opelz et al., 2006). Therefore, early detection and prediction of proteinuria are particularly important.

Nowadays, with significant improvements in reliability, sequencing chemistry, data interpretation, and costs, next-generation sequencing (NGS) has been increasingly applied in clinical practice, especially diagnostics. Based on the NGS techniques using massively parallel sequencing to decode large areas of the genome, we could potentially identify underlying causes of the disease, predict responses to interventions, and determine individuals at risk. To date, NGS has achieved great advances in infectious diseases (Gu et al., 2019), oncology (Kamps et al., 2017), and reproductive health (Mellis et al., 2018). Besides, Nagano C et al. found the common podocyte-related genes with mutations causing proteinuria were WT1, NPHS1, INF2, TRPC6, and LAMB2, by comprehensive gene screening of patients with diagnosis of nephrotic syndrome or glomerulosclerosis (Nagano et al., 2020). Bedin et al. (2019) reported 4 C-terminal CUBN variants are associated with proteinuria and slightly increased GFR, according to the NGS results from patients with suspected hereditary renal disease and chronic proteinuria. In our study, relying on the NGS technique, the TLR2 gene was detected associated with proteinuria after kidney transplantation. The *in vitro* experiment on podocytes further confirmed our finding.

TLRs are proven to be the human homolog of the *Drosophila* Toll protein. They have similar sequences and structures (Rock et al., 1998). Like *Drosophila* Toll, TLRs are type 1 membrane proteins with a conserved extracellular domain comprising leucine-rich repeats (LRRs) and a cytoplasmic domain related to the IL-1 receptor, termed the "Toll/IL-1 receptor (TIR) homology motif" (Medzhitov et al., 1997). The TLR family is a rapidly expanding family. So far, 11 human TLRs and 13 mouse TLR family members have been identified (O'Neill et al., 2013). Besides located in immunocytes, TLRs are distributed in various parenchymal cells. In the kidney, TLRs were found in both tubular cells and glomerular cells. Therefore, kidney diseases could be affected by the stimulation of TLRs on inflammatory cells or renal cells. Particularly, many experimental and emerging clinical data indicate that TLRs are involved in the pathogenesis of ischemia–reperfusion injury (I/R injury), urinary tract infections (UTIs), acute kidney injury (AKI), diabetic nephropathy, and lupus nephritis (LN).

TLR2 is one of the protein family members of TLRs. It forms TLR2 homodimers, TLR1–TLR2 heterodimers, or TLR2–TLR6 heterodimers to recognize plenty of ligands and triggers inflammatory response depending on the MyD88-signaling pathway. TLR2 can respond to various PAMPs and DAMPs (Shaw et al., 2011). For example, PGN, the main component of the Gram-positive bacteria wall, is considered the important ligand of TLR2. More and more evidence shows that many putative endogenous ligands have been discovered, which belong to the DAMP of TLR2. It is reported that TLR2 responds to endogenous ligands released during cellular stress and injury, such as lipopeptides, high mobility group box 1 (HMGB1), heat shock protein 60 (HSP60), and HSP70 (Asea et al., 2002; de Graaf et al., 2006). Recently, it has been proposed that biglycan, as an extracellular matrix breakdown product, can also activate TLR2. When TLR2 binds to ligands, it thereby activates NF-κB pathways

and finally induces proinflammatory cytokines, such as TNF-alpha and IL-6, from immune and non-immune cells. It was reported that TLR2 was expressed in kidney cells such as glomerular endothelial cells and renal tubular cells. It has been reported that proteinuria can cause renal tubular interstitial inflammation, accompanied by activation of the TLR2–MyD88–NF-κB pathway and secretion of proinflammatory cytokines TNF-alpha and IL-6 (Shigeoka et al., 2007). However, the specific mechanism of TLR2 gene acting on proteinuria after kidney transplantation still needs further study.

To sum up, there are many reasons for proteinuria after kidney transplantation, which have cross-effects on the damage of renal allograft function and are critical to the prognosis of the allograft. Monitoring proteinuria levels, and combining with the pathological examination of the transplanted kidney to early diagnose the cause of proteinuria, and corresponding treatment will effectively improve the long-term prognosis of the transplanted kidney. In this study, we indicated that the mutation of rs3804099 in the TLR2 gene was significantly related to the risk of proteinuria after kidney transplantation. The mutation of rs3804099 on the TLR2 gene may be related to podocyte injury and apoptosis. Besides, our data provide insights into the prediction of proteinuria and may imply potential individualized therapy for patients after kidney transplantation. However, limitations still exist in our study. First, three samples from each group were prepared for RNA sequencing, the number of which is too small. Additionally, normal kidney samples for RNA sequencing were excised far away from tumor tissue from patients undergoing radical nephrectomy. Hence, those are not complete normal samples. Besides, most of the patients involved in this study lacked the pathological results of the transplanted kidney biopsy, and the specific causes of proteinuria after kidney transplantation were not analyzed and grouped. Next, we plan to conduct analysis of the etiology of proteinuria to further clarify the roles and mechanisms of TLR2 in proteinuria after kidney transplantation.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The data presented in the study are deposited in the Sequence Read Archive (SRA) database repository, accession number (SRP133091). This data can be found here: [https://www.ncbi.nlm.nih.gov/sra/?term=SRP133091/SRP133091].

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the local ethics committee of the First Affiliated Hospital of Nanjing Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SF and ZG performed most of the experiments and interpreted the data. DF, ZW, MZ, and HC contributed to the analysis of the data. SF, ZG, and DF wrote the article. LS, JT, ZH, and XJ advised on the experimental design. MG, RT, and XL critically revised the manuscript and contributed to the conception and design. All authors read and approved the final article.

FUNDING

This work was supported by the National Natural Science Foundation of China (grant numbers 82070769, 81900684, 81870512, 81770751, 81570676), Project of Jiangsu Province for Important Medical Talent (grant number ZDRCA2016025), and "33 High Level Talents Project" in Jiangsu Province (grant numbers BRA2017532, BRA2016514, BRA2015469).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.798001/full#supplementary-material

REFERENCES

Asea, A., Rehli, M., Kabingu, E., Boch, J. A., Baré, O., Auron, P. E., et al. (2002). Novel Signal Transduction Pathway Utilized by Extracellular HSP70: role of toll-like receptor (TLR) 2 and TLR4 *J. Biol. Chem.* 277, 15028–15034. doi:10.1074/jbc.M200497200

Bedin, M., Boyer, O., Servais, A., Li, Y., Villoing-Gaudé, L., Tête, M.-J., et al. (2019). Human C-Terminal CUBN Variants Associate with Chronic Proteinuria and normal Renal Function. *J. Clin. Invest.* 130, 335–344. doi:10.1172/jci129937

Bergallo, M., Loiacono, E., Gambarino, S., Montanari, P., Galliano, I., and Coppo, R. (2017). Toll-like Receptors Are Essential for the Control of Endogenous Retrovirus Expression in Idiopathic Nephrotic Syndrome. *Minerva Urol. Nephrol.* 69, 201–208. doi:10.23736/s0393-2249.16.02658-8

Blasius, A. L., and Beutler, B. (2010). Intracellular Toll-like Receptors. *Immunity* 32, 305–315. doi:10.1016/j.immuni.2010.03.012

de Graaf, R., Kloppenburg, G., Kitslaar, P. J. H. M., Bruggeman, C. A., and Stassen, F. (2006). Human Heat Shock Protein 60 Stimulates Vascular Smooth Muscle Cell Proliferation through Toll-like Receptors 2 and 4. *Microbes Infect.* 8, 1859–1865. doi:10.1016/j.micinf.2006.02.024

Ding, L.-H., Liu, D., Xu, M., Wu, M., Liu, H., Tang, R.-N., et al. (2015). TLR2-MyD88-NF-κB Pathway Is Involved in Tubulointerstitial Inflammation Caused by Proteinuria. *Int. J. Biochem. Cel Biol.* 69, 114–120. doi:10.1016/j.biocel.2015.10.014

Fernandez-Fresnedo, G., Plaza, J. J., Sánchez-Plumed, J., Sanz-Guajardo, A., Palomar-Fontanet, R., and Arias, M. (2004). Proteinuria: a New Marker of Long-Term Graft and Patient Survival in Kidney Transplantation. *Nephrol. Dial. Transplant.* 19 (Suppl. 3), iii47–iii51. doi:10.1093/ndt/gfh1015

Gao, S.-L., Chen, Y.-D., Yue, C., Chen, J., Zhang, L.-F., Wang, S.-M., et al. (2019). -196 to -174del, rs4696480, rs3804099 Polymorphisms of Toll-like Receptor 2 Gene Impact the Susceptibility of Cancers: Evidence from 37053 Subjects. *Biosci. Rep.* 39, BSR20191698. doi:10.1042/bsr20191698

Gu, W., Miller, S., and Chiu, C. Y. (2019). Clinical Metagenomic Next-Generation Sequencing for Pathogen Detection. *Annu. Rev. Pathol. Mech. Dis.* 14, 319–338. doi:10.1146/annurev-pathmechdis-012418-012751

Harding, C. V., and Boom, W. H. (2010). Regulation of Antigen Presentation by *Mycobacterium tuberculosis*: a Role for Toll-like Receptors. *Nat. Rev. Microbiol.* 8, 296–307. doi:10.1038/nrmicro2321

Kamps, R., Brandão, R., Bosch, B., Paulussen, A., Xanthoulea, S., Blok, M., et al. (2017). Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *Ijms* 18, 308. doi:10.3390/ijms18020308

Kuliczkowska-Płaksej, J., Jończyk, M., Jawiarczyk-Przybyłowska, A., Stachowska, B., Zembska, A., Grzegrzółka, J., et al. (2021). The Frequency of TLR2 (Rs3804099, Rs3804100, and Rs5743708) and TLR4 (Rs4986790 and Rs4986791) Polymorphisms in Women with Polycystic Ovary Syndrome - Preliminary Study. *Gynecol. Endocrinol.* 37, 1027–1034. doi:10.1080/09513590.2021.1952975

Marshak-Rothstein, A. (2006). Toll-like Receptors in Systemic Autoimmune Disease. *Nat. Rev. Immunol.* 6, 823–835. doi:10.1038/nri1957

Medzhitov, R., Preston-Hurlburt, P., and Janeway, C. A., Jr. (1997). A Human Homologue of the Drosophila Toll Protein Signals Activation of Adaptive Immunity. *Nature* 388, 394–397. doi:10.1038/41131

Mellis, R., Chandler, N., and Chitty, L. S. (2018). Next-generation Sequencing and the Impact on Prenatal Diagnosis. *Expert Rev. Mol. Diagn.* 18, 689–699. doi:10.1080/14737159.2018.1493924

Menara, G., Lefort, N., Antignac, C., and Mollet, G. (2020). Generation of an Induced Pluripotent Stem Cell (iPSC) Line (IMAGINi007) from a Patient with Steroid-Resistant Nephrotic Syndrome Carrying the Homozygous p.R138Q Mutation in the Podocin-Encoding NPHS2 Gene. *Stem Cel Res.* 46, 101878. doi:10.1016/j.scr.2020.101878

Mirkamandar, E., Nemati, M., Hayatbakhsh, M. M., Bassagh, A., Khosravimashizi, A., and Jafarzadeh, A. (2018). Association of a Single Nucleotide Polymorphism in the TLR2 Gene (Rs3804099), but Not in the TLR4 Gene (Rs4986790), with *Helicobacter pylori* Infection and Peptic Ulcer. *Turk J. Gastroenterol.* 29, 283–291. doi:10.5152/tjg.2018.17484

Motojima, M., Matsusaka, T., Kon, V., and Ichikawa, I. (2010). Fibrinogen that Appears in Bowman's Space of Proteinuric Kidneys *In Vivo* Activates Podocyte Toll-like Receptors 2 and 4 *In Vitro*. *Nephron Exp. Nephrol.* 114, e39–e47. doi:10.1159/000254390

Nagano, C., Yamamura, T., Horinouchi, T., Aoto, Y., Ishiko, S., Sakakibara, N., et al. (2020). Comprehensive Genetic Diagnosis of Japanese Patients with Severe Proteinuria. *Sci. Rep.* 10, 270. doi:10.1038/s41598-019-57149-5

O'Neill, L. A. J., Golenbock, D., and Bowie, A. G. (2013). The History of Toll-like Receptors - Redefining Innate Immunity. *Nat. Rev. Immunol.* 13, 453–460. doi:10.1038/nri3446

Obhrai, J., and Goldstein, D. R. (2006). The Role of Toll-like Receptors in Solid Organ Transplantation. *Transplantation* 81, 497–502. doi:10.1097/01.tp.0000188124.42726.d8

Opelz, G., Zeier, M., Laux, G., Morath, C., and Döhler, B. (2006). No Improvement of Patient or Graft Survival in Transplant Recipients Treated with Angiotensin-Converting Enzyme Inhibitors or Angiotensin II Type 1 Receptor Blockers: a Collaborative Transplant Study Report. *Jasn* 17, 3257–3262. doi:10.1681/ASN.2006050543

Reichel, H., Zeier, M., and Ritz, E. (2004). Proteinuria after Renal Transplantation: Pathogenesis and Management. *Nephrol. Dial. Transplant.* 19, 301–305. doi:10.1093/ndt/gfh002

Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A., and Bazan, J. F. (1998). A Family of Human Receptors Structurally Related to Drosophila Toll. *Proc. Natl. Acad. Sci.* 95, 588–593. doi:10.1073/pnas.95.2.588

Rood, I. M., Deegens, J. K. J., Lugtenberg, D., Bongers, E. M. H. F., and Wetzels, J. F. M. (2019). Nephrotic Syndrome with Mutations in NPHS2: The Role of R229Q and Implications for Genetic Counseling. *Am. J. Kidney Dis.* 73, 400–403. doi:10.1053/j.ajkd.2018.06.034

Sancho, A., Gavela, E., Ávila, A., Morales, A., Fernández-Nájera, J. E., Crespo, J. F., et al. (2007). Risk Factors and Prognosis for Proteinuria in Renal Transplant Recipients. *Transplant. Proc.* 39, 2145–2147. doi:10.1016/j.transproceed.2007.07.005

Seki, E., and Brenner, D. A. (2008). Toll-like Receptors and Adaptor Molecules in Liver Disease: Update. *Hepatology* 48, 322–335. doi:10.1002/hep.22306

Shaw, J. L. V., Wills, G. S., Lee, K.-F., Horner, P. J., McClure, M. O., Abrahams, V. M., et al. (2011). *Chlamydia trachomatis* Infection Increases Fallopian Tube PROKR2 via TLR2 and NFκB Activation Resulting in a Microenvironment Predisposed to Ectopic Pregnancy. *Am. J. Pathol.* 178, 253–260. doi:10.1016/j.ajpath.2010.11.019

Shigeoka, A. A., Holscher, T. D., King, A. J., Hall, F. W., Kiosses, W. B., Tobias, P. S., et al. (2007). TLR2 Is Constitutively Expressed within the Kidney and Participates in Ischemic Renal Injury through Both MyD88-dependent and -independent Pathways. *J. Immunol.* 178, 6252–6258. doi:10.4049/jimmunol.178.10.6252

Ururahy, M. A. G., Loureiro, M. B., Freire-Neto, F. P., de Souza, K. S. C., Zuhl, I., Brandão-Neto, J., et al. (2012). Increased TLR2 Expression in Patients with Type 1 Diabetes: Evidenced Risk of Microalbuminuria. *Pediatr. Diabetes* 13, 147–154. doi:10.1111/j.1399-5448.2011.00794.x

Vachek, J., Maříková, A., Oulehle, K., Zakiyanov, O., and Tesař, V. (2020). Proteinuria from an Internists point of View. *Vnitr Lek* 66, 90–92. doi:10.36290/vnl.2020.052

Vilahur, G., and Badimon, L. (2014). Ischemia/reperfusion Activates Myocardial Innate Immune Response: the Key Role of the Toll-like Receptor. *Front. Physiol.* 5, 496. doi:10.3389/fphys.2014.00496

Wang, Z., Yang, H., Si, S., Han, Z., Tao, J., Chen, H., et al. (2018). Polymorphisms of Nucleotide Factor of Activated T Cells Cytoplasmic 2 and 4 and the Risk of Acute Rejection Following Kidney Transplantation. *World J. Urol.* 36, 111–116. doi:10.1007/s00345-017-2117-2

Wang, Z., Zhang, H., Yang, H., Zheng, M., Guo, M., Chen, H., et al. (2020). An Intronic Polymorphism of NFATC1 Gene Shows a Risk Association with Biopsy-Proven Acute Rejection in Renal Transplant Recipients. *Ann. Transl Med.* 8, 211. doi:10.21037/atm.2020.01.61

Conflict of Interest: RNA sequencing was performed by Nanjing Jixu Zhigu Biotechnology Co., LTD.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Genetics

ORIGINAL RESEARCH
published: 18 March 2022
doi: 10.3389/fgene.2022.810292

Identification of *GGT5* as a Novel Prognostic Biomarker for Gastric Cancer and its Correlation With Immune Cell Infiltration

Yuli Wang[1†], *Yuan Fang*[1†], *Fanchen Zhao*[1], *Jiefei Gu*[2], *Xiang Lv*[1], *Rongzhong Xu*[1], *Bo Zhang*[1], *Zhihong Fang*[1*] and *Yan Li*[3*]

[1]*Department of Oncology II, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China,* [2]*Information Center, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China,* [3]*Department of Oncology I, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China*

OPEN ACCESS

Edited by:
Yuriy L. Orlov,
I. M. Sechenov First Moscow State Medical University, Russia

Reviewed by:
Elvira Galieva,
Novosibirsk State University, Russia
Maxim Sorokin,
I. M. Sechenov First Moscow State Medical University, Russia
Noha Said Helal,
Theodor Bilharz Research Institute, Egypt

***Correspondence:**
Zhihong Fang
fangzhihong@shutcm.edu.cn
Yan Li
yan.xiaotian@shutcm.edu.cn

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *06 November 2021*
Accepted: *03 March 2022*
Published: *18 March 2022*

Citation:
Wang Y, Fang Y, Zhao F, Gu J, Lv X, Xu R, Zhang B, Fang Z and Li Y (2022) Identification of GGT5 as a Novel Prognostic Biomarker for Gastric Cancer and its Correlation With Immune Cell Infiltration. Front. Genet. 13:810292. doi: 10.3389/fgene.2022.810292

Gastric cancer (GC) is a common malignant tumor of the digestive system. Recent studies revealed that high gamma-glutamyl-transferase 5 (*GGT5*) expression was associated with a poor prognosis of gastric cancer patients. In the present study, we aimed to confirm the expression and prognostic value of *GGT5* and its correlation with immune cell infiltration in gastric cancer. First, we compared the differential expression of *GGT5* between gastric cancer tissues and normal gastric mucosa in the cancer genome atlas (TCGA) and GEO NCBI databases using the most widely available data. Then, the Kaplan-Meier method, Cox regression, and univariate logistic regression were applied to explore the relationships between *GGT5* and clinical characteristics. We also investigated the correlation of *GGT5* with immune cell infiltration, immune-related genes, and immune checkpoint genes. Finally, we estimated enrichment of gene ontologies categories and relevant signaling pathways using GO annotations, KEGG, and GSEA pathway data. The results showed that *GGT5* was upregulated in gastric cancer tissues compared to normal tissues. High *GGT5* expression was significantly associated with T stage, histological type, and histologic grade ($p < 0.05$). Moreover, gastric cancer patients with high *GGT5* expression showed worse 10-years overall survival ($p = 0.008$) and progression-free intervals ($p = 0.006$) than those with low *GGT5* expression. Multivariate analysis suggested that high expression of *GGT5* was an independent risk factor related to the worse overall survival of gastric cancer patients. A nomogram model for predicting the overall survival of GC was constructed and computationally validated. *GGT5* expression was positively correlated with the infiltration of natural killer cells, macrophages, and dendritic cells but negatively correlated with Th17 infiltration. Additionally, we found that *GGT5* was positively co-expressed with immune-related genes and immune checkpoint genes. Functional analysis revealed that differentially expressed genes relative to *GGT5* were mainly involved in the biological processes of immune and inflammatory responses. In conclusion, *GGT5* may serve as a promising prognostic biomarker and a potential immunological therapeutic target for GC, since it is associated with immune cell infiltration in the tumor microenvironment.

Keywords: bioinformatics, GGT5, gastric cancer, immune infiltration, prognostic biomarker, TCGA database, gene expression

INTRODUCTION

Gastric cancer (GC) is one of the most common digestive malignancies, ranking fifth in tumor morbidity and fourth in mortality worldwide (Sung et al., 2021). In recent years, with the increasing popularity of screening gastroscopy and surgical intervention, there has been a decline in the incidence and mortality of non-cardia gastric cancer. However, the progressively increased incidence of early-onset GC with more aggressive features is garnering attention and warrants deeper investigation (Bergquist et al., 2019). Since inhibitors of human epidermal growth factor receptor 2 (HER2), such as trastuzumab, pertuzumab, and lapatinib, were introduced for the treatment of HER2-overexpressing GC, the outcome of advanced gastric cancer patients has been significantly improved (Bang et al., 2010). Nevertheless, previous studies have indicated that the majority of gastric cancer patients suffer acquired resistance to trastuzumab within a relatively short period (Zhang et al., 2019). In addition, GC is a highly heterogeneous tumor, and HER2 positivity is found in only approximately 13.0%–22.0% of GC cases (Cappellesso et al., 2015; Van Cutsem et al., 2015; Abrahao-Machado and Scapulatempo-Neto, 2016). There is still a lack of specific biomarkers for early diagnosis or use as potential therapeutic targets for GC. Hence, there is a great need to identify potential prognostic biomarkers or therapeutic targets to improve the survival of gastric cancer patients.

As a crucial liver enzyme involved in extracellular glutathione metabolism, gamma-glutamyl transferase can cleave glutathione peptides to maintain the glutathione balance in the human body (Heisterkamp et al., 2008). Gamma-glutamyltransferase 5 (*GGT5*) is one of the two GGT family members (*GGT1* and *GGT5*) with catalytic activity identified to date and was first reported in detail in 2008 (Heisterkamp et al., 2008). *GGT5* is widely distributed in a variety of tissues, with relatively high expression in liver, kidney, and alveolar macrophages (Hanigan et al., 2015). Functional analyses showed that *GGT5* played a pivotal role in oxidative regulation, drug metabolism, and immune modulation in the human body (Wickham et al., 2011).

Recent studies have demonstrated that upregulated *GGT5* is correlated with tumorigenesis and the progression of a variety of malignancies, including GC (Ren et al., 2020; Wei et al., 2020; Wen et al., 2020; Wu et al., 2021; Ye et al., 2021). Three bioinformatics studies based on the clinical gastric cancer samples retrieved from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases indicated that *GGT5* was included in prognostic gene signatures, and overexpression of *GGT5* was inversely correlated with the survival in GC patients (Wei et al., 2020; Wen et al., 2020; Ye et al., 2021). It was presumed that the underlying mechanism by which *GGT5* affects the prognosis of GC patients might be associated with metabolic regulation, immune modulation, and antioxidant effects (Wei et al., 2020; Wen et al., 2020; Ye et al., 2021). These findings indicated that *GGT5* might serve as a promising prognostic biomarker or potential therapeutic target for gastric cancer. However, the correlation of *GGT5* expression levels with immune cell infiltration in the tumor microenvironment of gastric cancer has not yet been investigated yet.

In the present study, integrated bioinformatics analysis was carried out based on RNA sequencing data retrieved from the TCGA (https://www.cancer.gov/tcga/) and validated in the GEO (https://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2012) database. We first compared the differential expression levels of *GGT5* between GC tissues and normal gastric mucosa in the TCGA database and simultaneously validated them in the other two independent RNA profiles, GSE54129 (https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE54129) and GSE29272 (Wang et al., 2013; Cheng et al., 2017) from the GEO database. Subsequently, we identified the differentially expressed genes (DEGs) between the *GGT5*-high and *GGT5*-low expression groups. The Kaplan-Meier method, Cox regression, and univariate logistic regression (Zhu C. et al., 2021) were performed to investigate the relationships between *GGT5* and clinical characteristics in gastric cancer patients. A nomogram model for predicting overall survival was constructed and computationally validated. Moreover, single-sample gene set enrichment analysis (ssGSEA) (Bindea et al., 2013) was employed to analyze the correlation of *GGT5* expression with infiltration patterns for 24 immune cell types in GC samples. We also explored the correlation of *GGT5* expression with immune-related genes and immune checkpoint genes to further understand the underlying mechanism by which *GGT5* is correlated with immune cell infiltration in gastric cancer. Finally, gene ontology annotation, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA) (Chen et al., 2022) were applied to explore the potential functions of *GGT5* in gastric cancer.

MATERIALS AND METHODS

Data Acquisition and Preprocessing

High-throughput sequencing data (HTSeq) with clinical information, including 375 GC tissues and 32 adjacent normal tissues were downloaded from the TCGA database (https://www.cancer.gov/tcga/). The Fragments Per kilobase Million (FPKM) data were quantified in transcripts per million (TPM). All FPKM values were then log-transformed to obtain a normal distribution with log2(FPKM+1) for further statistical analyses. Similarly, the raw values of the microarray expression data were downloaded from the GEO database. Two datasets, GSE54129 and GSE29272, were ultimately screened out according to the following inclusion criteria: (1) achievable comparison of gastric cancer tissues with normal gastric mucosa limited to *Homo sapiens*; (2) no medical intervention (chemotherapy, radiotherapy, and/or targeted therapy) before sample collection; and (3) more than 100

samples in the full study, and the raw gene expression data can be downloaded in CEL format for further analysis. Subsequently, the data were log2 transformed and quantile normalized using R software (version 3.6.3). The probe ID for each gene was then converted into a gene symbol, and the average expression value was taken while multiple probes were converted to one gene symbol. Moreover, the batch effects were corrected by applying the "remove Batch Effect" function in the "limma" package of R software.

Differential Expression Analysis of *GGT5* in the TCGA and GEO Databases

We compared the expression level of *GGT5* across cancers and corresponding normal tissues using the Wilcoxon rank-sum test based on the TCGA Pan-Cancer dataset (https://gdc.cancer.gov/about-data/publications/pancanatlas). For gastric cancer, the difference in *GGT5* between tumor tissues and normal tissues was evaluated in TCGA and the Human Protein Atlas (HPA) (Access date: August 2021, https://www.proteinatlas.org/) databases. Additionally, we utilized the area under the ROC curve using R software (version 3.6.3) to assess the predictive value of *GGT5* in distinguishing gastric cancer tissues and normal gastric tissues. Two RNA expression datasets, GSE54129 (https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE54129) and GSE29272 (Wang et al., 2013; Cheng et al., 2017), retrieved from the GEO database were used as the external validation cohorts.

Identification of Differentially Expressed Genes Relative to *GGT5*

Gastric cancer tissue samples were classified into high and low expression groups by median value splitting. The identification of DEGs between the *GGT5*-high and *GGT5*-low expression groups was performed by Wald's test with the R package DESeq2 (version 1.26.0) (Love et al., 2014). The screening threshold for statistical significance was set as absolute log2-fold change (FC) > 2 and adjusted p *value* < 0.05. The upregulated and downregulated DEGs were visualized using volcano plots, and *GGT5*-related DEGs were displayed in a heatmap plot.

The Relationship Between *GGT5* Expression and the Clinical Characteristics of Patients With Gastric Cancer

First, we determined the cut-off of *GGT5* expression according to its median value. To investigate the relationship between *GGT5* expression and the clinical characteristics of gastric cancer patients, we used univariate logistic regression analysis. The 10-years overall survival (OS), progression-free interval (PFI), and disease-specific survival (DSS) in the *GGT5*-high and *GGT5*-low groups were compared using the Kaplan–Meier curve and the log-rank test (Huang X. et al., 2021). Univariate and multivariate Cox regression analyses based on *GGT5* expression and clinical characteristics were applied to screen the independent survival risk factors for gastric cancer. Additionally, subgroup analysis was conducted and displayed in the forest plots by R software (version 3.6.3). The forest plots show the hazard ratio (HR) and 95% confidence interval of the prognostic factors by univariate and multivariate analyses. All p *value*s less than 0.05 were considered statistically significant.

Construction and Evaluation of a Prognostic Model for Gastric Cancer Patients

Based on the results of univariate and multivariate Cox regression analyses, prognostic factors were included to build a nomogram model, which aimed to predict the overall survival of gastric cancer patients at 1, 2, and 3 years. The nomogram analysis was carried out using the "RMS" package in R software (version 3.6.3). The nomogram was assessed graphically by plotting the calibration curves, which compared the observed values (Kaplan-Meier method) with the nomogram-predicted probabilities. For a well-calibrated model, the scatter points of the nomogram prediction model will fall on a 45-degree diagonal line (Balachandran et al., 2015). Furthermore, the Harrell concordance index (C-index) was also employed to evaluate the overall predictive ability of the nomogram model. The value of the C-index ranges from 0.5 to 1, and the higher the C-index, the better the prediction performance. The significance level for this study was set at 0.05, and all statistical tests were two-tailed.

The Correlation of *GGT5* Expression With Immune Cell Infiltration and Immune-Related Genes

To explore the associations between *GGT5* expression and immune cell infiltration, we applied the ssGSEA algorithm using the R package 'GSVA' (version 1.34.0) (Bindea et al., 2013) to investigate the immune infiltration landscape of 24 different immune cells in both the *GGT5*-high and *GGT5*-low groups. The Pearson correlation coefficients were calculated to determine the relationship between *GGT5* expression and infiltrating immune cells. Representative immune-infiltrating cells with significant correlations are displayed in scatter plots and column bar graphs. Moreover, to further understand the potential mechanisms by which *GGT5* is correlated with immune cell infiltration, we also assessed the correlation of *GGT5* expression with immune-related genes, including MHC genes, immune activation genes, immunosuppressive genes, chemokine receptors, and chemokines (Zhu H. et al., 2021).

Immune Checkpoint Analysis

The expression level of immune checkpoint biomarkers was closely related to the therapeutic response to immune checkpoint blockade treatment. Therefore, the present study focused on eight critical immune checkpoint genes, including *SIGLEC15*, *TIGIT*, *CD274*, *HAVCR2*, *PDCD1*, *CTLA4*, *LAG3*, and *PDCD1LG2*. Different expression levels of these immune checkpoints were compared between the *GGT5*-high and *GGT5*-low expression groups by heatmap and bar charts. We also applied the R package "immuneeconv" to assess the co-expression of *GGT5* with immune checkpoint biomarkers based on the TCGA database (https://www.cancer.gov/tcga/). Additionally, the TIDE (Jiang et al., 2018) algorithm was employed to predict the potential response to immune checkpoint blockade therapy.

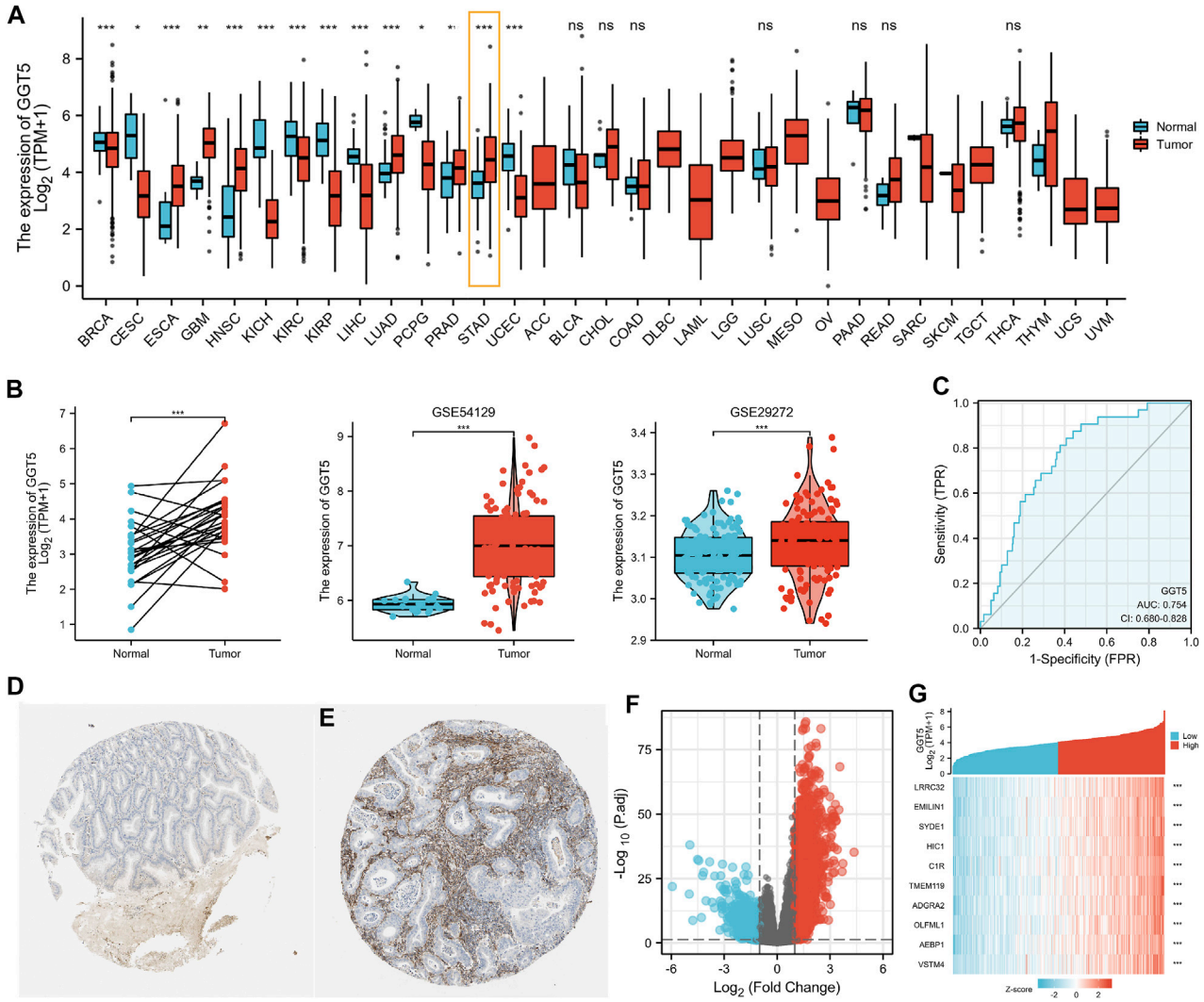


FIGURE 1 | Pan-cancer analysis of *GGT5* expression levels and upregulated *GGT5* in human gastric cancer. **(A)** Detection of the expression level of *GGT5* in a pan-cancer dataset from TCGA. **(B)** The expression level of *GGT5* in gastric cancer compared to paired and normal gastric mucosa from the GEO (GSE54129, GSE29272) and TCGA datasets. **(C)** ROC curve showing the predictive value of *GGT5* for identifying gastric cancer tissues based on the TCGA database. **(D,E)** Comparison of *GGT5* between normal gastric mucosa **(D)** and gastric cancer tissue **(E)** based on the Human Protein Atlas database (antibody HPA008121, 10×) (accession date: August 2021, https://www.proteinatlas.org/). **(F)** The volcano plot presents the *GGT5*-related differentially expressed genes (DEGs) based on the TCGA database. Red and blue dots represent up and downregulated genes, respectively. **(G)** The heatmap shows the top 10 DEGs related to *GGT5*. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ns, not significant, $p > 0.05$. ACC: adrenocortical carcinoma; BLCA: bladder urothelial carcinoma; BRCA: breast invasive carcinoma; CESC: cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL: cholangiocarcinoma; COAD: colon adenocarcinoma; DLBC: lymphoid neoplasm diffuse large B-cell lymphoma; ESCA: esophageal carcinoma; GBM: glioblastoma multiforme; HNSC: head and neck squamous cell carcinoma; KICH: kidney chromophobe; KIRC: kidney renal clear cell carcinoma; KIRP: kidney renal papillary cell carcinoma; LAML: acute myeloid leukemia; LGG: brain lower grade glioma; LIHC: liver hepatocellular carcinoma; LUAD: lung adenocarcinoma; LUSC: lung squamous cell carcinoma; MESO: mesothelioma; OV: ovarian serous cystadenocarcinoma; PAAD: pancreatic adenocarcinoma; PCPG: pheochromocytoma and paraganglioma; PRAD: prostate adenocarcinoma; READ: rectum adenocarcinoma; SARC: sarcoma; SKCM: skin cutaneous melanoma; STAD: stomach adenocarcinoma; TGCT: testicular germ cell tumors; THCA: thyroid carcinoma; THYM: thymoma; UCEC: uterine corpus endometrial carcinoma; UCS: uterine carcinosarcoma; UVM: uveal melanoma.

Functional Analyses of *GGT5* in Gastric Cancer

Before performing gene functional enrichment analysis, we transformed the gene symbols into EnterzID and obtained GO and KEGG signaling pathway annotations with R software. To further understand the functions of *GGT5* in gastric cancer, the clusterProfiler (version 3.14.3) package (Yu et al., 2012) was utilized to carry out GO functional enrichment analysis and it displayed the three aspects of GO enrichment, including biological processes, cellular components, and molecular functions. Enrichment analysis of DEGs was also conducted with the clusterProfiler package, and the significant DEG-

related signaling pathways were mapped into a bubble graph. For GSEA, based on the mean expression of *GGT5*, the candidate genes were divided into the *GGT5*-high group and the *GGT5*-low group. Functional predefined gene sets were obtained from the Molecular Signatures database, MSigDB (https://www.gsea-msigdb.org/gsea/msigdb). The candidate genes involved in the pathway with the screening criteria of $p < 0.05$ and false discovery rate (FDR) < 0.25 were considered significantly enriched. The normalized enrichment score and adjusted *p value* were applied to select the significantly enriched signaling pathways.

RESULTS

The Expression Level of *GGT5* was Elevated in Human Gastric Tissues

We first investigated *GGT5* expression across cancers in a pan-cancer dataset from TCGA. The results indicated that the expression level of *GGT5* was significantly elevated in esophageal carcinoma (ESCA), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), lung adenocarcinoma (LUAD), prostate adenocarcinoma (PRAD), and stomach adenocarcinoma (STAD) ($p < 0.05$). In contrast, significantly decreased *GGT5* expression was noticed in invasive breast carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), and uterine corpus endometrial carcinoma (UCEC) ($p < 0.05$) (**Figure 1A**). There was a large degree of heterogeneity across cancers, and only a few of the tumor types, including gastric cancer, had high *GGT5* expression levels.

Next, we compared the expression level of *GGT5* between gastric cancer tissues and adjacent normal tissues using the TCGA database and further validated it using related array data (GSE54129, GSE29272) from the GEO database. All of the results indicated that *GGT5* expression levels were elevated in gastric cancer tissues compared with normal tissues ($p < 0.05$) (**Figure 1B**, **Supplementary Table S1**). The ROC curve was plotted to determine the sensitivity and specificity of *GGT5* to distinguish gastric cancer tissues from normal gastric mucosa. The results showed that the area under the ROC curve was 0.754, which indicated that *GGT5* might potentially contribute to the identification of gastric cancer tissues (**Figure 1C**). Similar results were also obtained in the validation cohorts of two GEO datasets, as shown in **Supplementary Figures S1A, S2A**. Furthermore, the protein level of *GGT5* was also found to be higher in gastric cancer tissue than that in normal tissues in the HPA database (**Figures 1D,E**).

Identification of Differentially Expressed Genes in Gastric Cancer

According to the gene expression level of *GGT5*, the TCGA stomach adenocarcinoma samples (https://www.cancer.gov/tcga/) were stratified into a high-expression group ($N = 203$) and a low-expression group ($N = 204$) with a total number of uni-genes of 56,493 (**Supplementary Table S2**). Finally, a total of 330 DEGs were screened out with the criteria of $|\log2(FC)|>2$ and adjusted *p value* < 0.05, including 216 upregulated genes and 114 downregulated genes (**Figure 1F**, **Supplementary Table S3**). The HTSeq-Count data (Wang et al., 2013; Cheng et al., 2017) were further analyzed using the DESeq2 package in R software. The heatmap showed the top 10 DEGs that were most closely related to *GGT5* expression (**Figure 1G**). Similarly, 507 DEGs, including 304 upregulated genes and 203 downregulated genes were identified from GSE54129 (**Supplementary Figures S1B,C** and **Supplementary Table S4**), and a total of 38 DEGs, including 37 upregulated genes and one downregulated gene, were detected from GSE29272 (**Supplementary Figures S2B,C** and **Supplementary Table S5**).

Correlation Between *GGT5* Expression Level and the Clinicopathological Features of Gastric Cancer Patients

To define the clinical correlation between the *GGT5* expression level and the clinicopathological features in gastric cancer, we further analyzed the differences in clinical characteristics between the *GGT5*-high and *GGT5*-low expression groups based on the TCGA database. After removing duplicates, a total of 375 patients, including 241 (64.3%) men and 134 (35.7%) women, were included in the analysis, as displayed in **Table 1**. Our results revealed that *GGT5* overexpression was closely correlated with the pathological stage (Stage II & Stage III & Stage IV vs. Stage I, $p < 0.05$), T stage (T2–T4 vs. T1, $p < 0.001$), histological type (diffuse type and mucinous type and signet ring type vs. tubular type, $p < 0.05$), overall survival status (alive vs. dead, $p < 0.01$), histologic grade (G3 vs. G1 & G2, $p < 0.001$), and age (less than or equal to 60 vs. greater than 60, $p < 0.01$) (**Figures 2A–F**). Nevertheless, the expression level of *GGT5* had no significant correlation with M stage, N stage, or residual tumor ($p > 0.05$). Furthermore, the univariate logistic regression analysis indicated that the expression level of *GGT5* was closely associated with the clinical characteristics of a poor prognosis in patients with gastric cancer (**Table 2**). Specifically, *GGT5* was positively correlated with the T stage (T3&T4 vs. T1&T2, OR = 1.437, 95% CI: 1.167–1.783, $p < 0.001$), histologic grade (G3 vs. G1&G2, OR = 1.738, 95% CI: 1.420–2.151, $p < 0.001$), and histological type (tubular type vs. not otherwise specified, OR = 0.658, 95% CI: 0.505–0.847, $p = 0.001$). Collectively, *GGT5*-high expression gastric cancers were associated with a relatively higher pathological stage (TNM stage and T stage) and histological grade. All of the above results suggested that upregulated *GGT5* was strongly correlated with a poor tumor differentiation grade and a poor prognosis of gastric cancer.

GGT5 Overexpression was Found to be Associated With Poor Outcomes in Gastric Cancer Patients

To evaluate the prognostic value of *GGT5* in gastric cancer, we calculated the survival of patients with different *GGT5* expression levels using the Kaplan-Meier method. The results indicated that

TABLE 1 | Clinical characteristics of gastric cancer patients in the *GGT5*-high and the *GGT5*-low expression groups.

Characteristic	Levels	Low expression of *GGT5*	High expression of *GGT5*	*p*	Method
n		187	188		
T stage, n (%)	T1	17 (4.6%)	2 (0.5%)	<0.001	Chi-square
	T2	46 (12.5%)	34 (9.3%)		
	T3	75 (20.4%)	93 (25.3%)		
	T4	45 (12.3%)	55 (15.0%)		
N stage, n (%)	N0	59 (16.5%)	52 (14.6%)	0.135	Chi-square
	N1	50 (14.0%)	47 (13.2%)		
	N2	41 (11.5%)	34 (9.5%)		
	N3	28 (7.8%)	46 (12.9%)		
M stage, n (%)	M0	166 (46.8%)	164 (46.2%)	1.000	Chi-square
	M1	13 (3.7%)	12 (3.4%)		
Pathologic stage, n (%)	Stage I	38 (10.8%)	15 (4.3%)	0.003	Chi-square
	Stage II	50 (14.2%)	61 (17.3%)		
	Stage III	67 (19.0%)	83 (23.6%)		
	Stage IV	22 (6.2%)	16 (4.5%)		
Primary therapy outcome, n (%)	PD	33 (10.4%)	32 (10.1%)	0.764	Fisher's test
	SD	7 (2.2%)	10 (3.2%)		
	PR	1 (0.3%)	3 (0.9%)		
	CR	111 (35.0%)	120 (37.9%)		
Gender, n (%)	Female	66 (17.6%)	68 (18.1%)	0.945	Chi-square
	Male	121 (32.3%)	120 (32.0%)		
Race, n (%)	Asian	43 (13.3%)	31 (9.6%)	<0.001	Chi-square
	Black or African American	10 (3.1%)	1 (0.3%)		
	White	97 (30.0%)	141 (43.7%)		
Age, n (%)	≤65	74 (19.9%)	90 (24.3%)	0.153	Chi-square
	>65	110 (29.6%)	97 (26.1%)		
Histological type, n (%)	Diffuse Type	26 (7.0%)	37 (9.9%)	0.005	Chi-square
	Mucinous Type	4 (1.1%)	15 (4.0%)		
	Not Otherwise, Specified	103 (27.5%)	104 (27.8%)		
	Papillary Type	4 (1.1%)	1 (0.3%)		
	Signet Ring Type	5 (1.3%)	6 (1.6%)		
	Tubular Type	45 (12.0%)	24 (6.4%)		
Residual tumor, n (%)	R0	150 (45.6%)	148 (45.0%)	0.262	Chi-square
	R1	5 (1.5%)	10 (3.0%)		
	R2	10 (3.0%)	6 (1.8%)		
Histologic grade, n (%)	G1	6 (1.6%)	4 (1.1%)	<0.001	Fisher's test
	G2	89 (24.3%)	48 (13.1%)		
	G3	87 (23.8%)	132 (36.1%)		
Anatomic neoplasm subdivision, n (%)	Antrum/Distal	73 (20.2%)	65 (18.0%)	0.054	Fisher's test
	Cardia/Proximal	22 (6.1%)	26 (7.2%)		
	Fundus/Body	56 (15.5%)	74 (20.5%)		
	Gastroesophageal Junction	25 (6.9%)	16 (4.4%)		
	Other	4 (1.1%)	0 (0%)		
Anti-reflux treatment, n (%)	No	79 (44.1%)	63 (35.2%)	0.386	Chi-square
	Yes	17 (9.5%)	20 (11.2%)		
Reflux history, n (%)	No	92 (43.0%)	83 (38.8%)	0.258	Chi-square
	Yes	25 (11.7%)	14 (6.5%)		
H pylori infection, n (%)	No	90 (55.2%)	55 (33.7%)	0.464	Chi-square
	Yes	9 (5.5%)	9 (5.5%)		
Barrett's esophagus, n (%)	No	111 (53.4%)	82 (39.4%)	0.152	Chi-square
	Yes	12 (5.8%)	3 (1.4%)		
OS event, n (%)	Alive	122 (32.5%)	106 (28.3%)	0.099	Chi-square
	Dead	65 (17.3%)	82 (21.9%)		
DSS event, n (%)	Alive	135 (38.1%)	128 (36.2%)	0.203	Chi-square
	Dead	39 (11.0%)	52 (14.7%)		
PFI event, n (%)	Alive	136 (36.3%)	115 (30.7%)	0.023	Chi-square
	Dead	51 (13.6%)	73 (19.5%)		
Age, median (IQR)		68 (58.75, 74)	66 (57.50, 72)	0.118	Wilcoxon

the 10-years overall survival of patients with low *GGT5* expression was better than that of patients with high *GGT5* expression (HR = 1.58, 95% CI: 1.130–2.200, p = 0.008, **Figure 3A**). Similar results were also obtained for the progression-free interval (PFI, HR = 1.67, 95% CI: 1.160–2.390, p = 0.006, **Figure 3B**). Patients with high *GGT5*

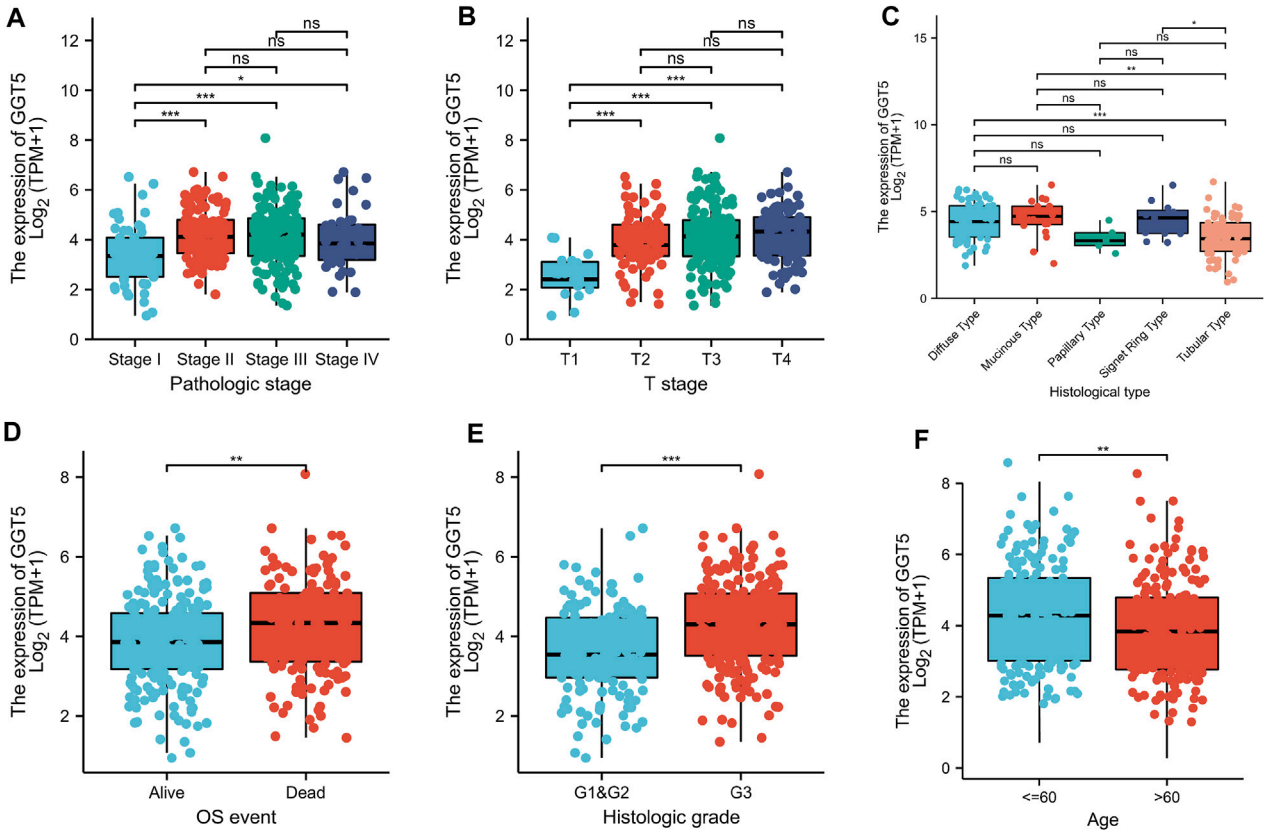


FIGURE 2 | Correlation between *GGT5* expression level and the clinicopathological features of gastric cancer patients for **(A)** pathologic stage, **(B)** T stage, **(C)** histological type, **(D)** OS event, **(E)** histologic grade, and **(F)** age. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ns, not significant, $p > 0.05$.

TABLE 2 | Univariate logistic regression assesses the relationships between *GGT5* and clinical characteristics of gastric cancer patients.

Characteristics	Total (*n*)	Odds ratio (OR)	*p*- value
T stage (T3&T4 vs. T1&T2)	367	1.437 (1.167–1.783)	<0.001
Histological type (Tubular Type vs. Not Otherwise, Specified)	276	0.658 (0.505–0.847)	0.001
Histologic grade (G3 vs. G1&G2)	366	1.738 (1.420–2.151)	<0.001

expression showed a similar prolonged survival trend for disease-specific survival (DSS, HR = 1.52, 95% CI: 0.99–2.32, $p = 0.054$, **Figure 3C**), but the difference was not statistically significant. Furthermore, the subgroup analysis demonstrated that the overall survival of patients with *GGT5*-high expression showed a significantly poor prognosis in age ≤65, pathologic stage III, T3 & T4 stage, N1 stage, and M0 stage (**Figures 3D–H**).

Univariate Cox regression analysis revealed that T stage, N stage, M stage, pathologic stage, primary therapy outcome, residual tumor, and age were closely correlated with the overall survival of the GC patients, as shown in **Table 3** and **Figure 4A**. However, the prognostic value of *GGT5* for GC patients was not statistically significant (HR = 1.330, 95% CI: 0.956–1.851, $p = 0.091$). Further multivariate analyses were carried out to screen independent prognostic factors. The results showed that in addition to the previously mentioned primary therapy outcome (PD & SD & PR vs. CR, HR = 4.528, 95% CI: 2.885–7.107, $p < 0.001$) and age (>65 vs. ≤65, HR = 1.744, 95% CI: 1.121–2.712, $p = 0.014$), high *GGT5* expression was also an independent prognostic factor for worse overall survival in gastric cancer patients (HR = 1.724, 95% CI: 1.094–2.717, $p = 0.019$) (**Table 3** and **Figure 4B**). To summarize, our study revealed that upregulated expression of *GGT5* was correlated with shorter overall survival and progression-free interval of GC patients.

Construction and Computational Validation of a Nomogram Model for Gastric Cancer Patients Based on *GGT5*

To establish a novel model for predicting the outcomes of patients with gastric cancer, a nomogram was constructed for overall

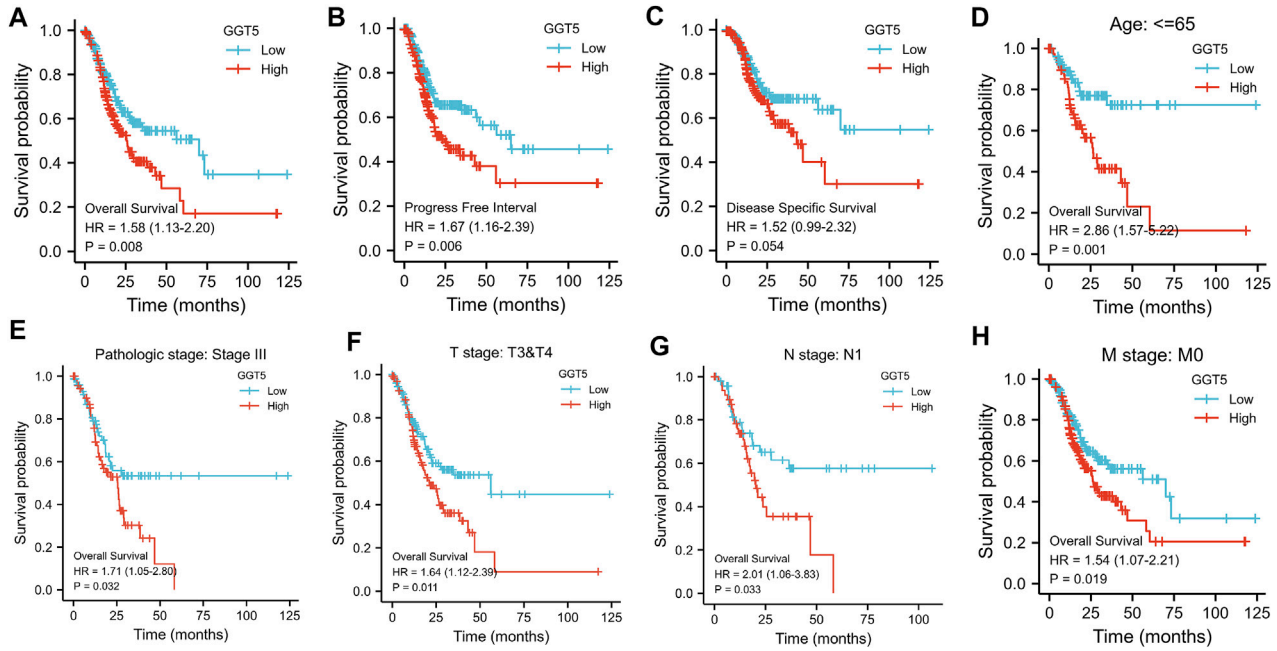


FIGURE 3 | Kaplan-Meier analysis of gastric cancer patients according to *GGT5* expression level and subgroup analysis. The overall survival curves **(A),** progression-free interval curves **(B)**, and disease-specific survival curves **(C)** between the *GGT5*-high and *GGT5*-low expression groups. **(D–H)** Comparison of the overall survival curves in age ≤65, pathologic stage III, T3 & T4 stage, N1 stage, and M0 stage subgroups between the *GGT5*-high and *GGT5*-low expression groups.

TABLE 3 | Univariate and multivariate Cox regression analysis of the risk factors for OS in patients with gastric cancer.

Characteristics	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	*p value*	Hazard ratio (95% CI)	*p value*
T stage (T3&T4 vs. T1&T2)	362	1.719 (1.131–2.612)	**0.011**	1.109 (0.599–2.054)	>0.05
N stage (N1&N2&N3 vs. N0)	352	1.925 (1.264–2.931)	**0.002**	1.483 (0.705–3.120)	>0.05
M stage (M1 vs. M0)	352	2.254 (1.295–3.924)	**0.004**	1.078 (0.428–2.711)	>0.05
Pathologic stage (III & IV vs. I & II)	347	1.947 (1.358–2.793)	**<0.001**	1.245 (0.654–2.370)	>0.05
Primary outcome (PD & SD & PR vs. CR)	313	4.228 (2.905–6.152)	**<0.001**	4.528 (2.885–7.107)	**<0.001**
Residual tumor (R1 & R2 vs. R0)	325	3.445 (2.160–5.494)	**<0.001**	1.140 (0.600–2.164)	>0.05
Age (>65 vs. ≤65)	367	1.620 (1.154–2.276)	**0.005**	1.744 (1.121–2.712)	**0.014**
GGT5 (High vs. Low)	370	1.330 (0.956–1.851)	>0.05	1.724 (1.094–2.717)	**0.019**

Statistical p *values < 0.05 are shown in bold.*

survival (**Figure 4C**). The variables of the nomogram were selected according to the results of the univariate and multivariate regression, and a 100-point scale was used to assign a point value to each variable. Each point of the variable was summed, and the total prone score was calculated, ranging from 0 to 240 points. Subsequently, by drawing a vertical line from the total points scale to the survival probability lines, we could obtain the estimated probabilities of 1-year, 2-years, and 3-years overall survival for GC patients.

To evaluate the accuracy and reliability of the nomogram model in predicting survival, the C-index and calibration curve were both used for further computational validation. The results revealed that the C-index of the nomogram model was 0.724 (95% CI: 0.698–0.749), which implied that the novel model was moderately accurate and appropriate for the overall survival prediction of GC patients. Additionally, it was found that the bias-corrected lines of 1 year, 2 years, and 3 years were close to the ideal 45° diagonal line in the calibration plot, which indicated that the theoretical values were in agreement with the observed values (**Figure 4D**). The above outcomes confirmed that the nomogram model could be applied for predicting the 1-year, 2-years, and 3-years overall survival of patients with gastric cancer.

Relationship Between *GGT5* Expression and Immune Cell Infiltration in Gastric Cancer

To investigate the association between the expression level of *GGT5* and immune infiltration, we first compared the

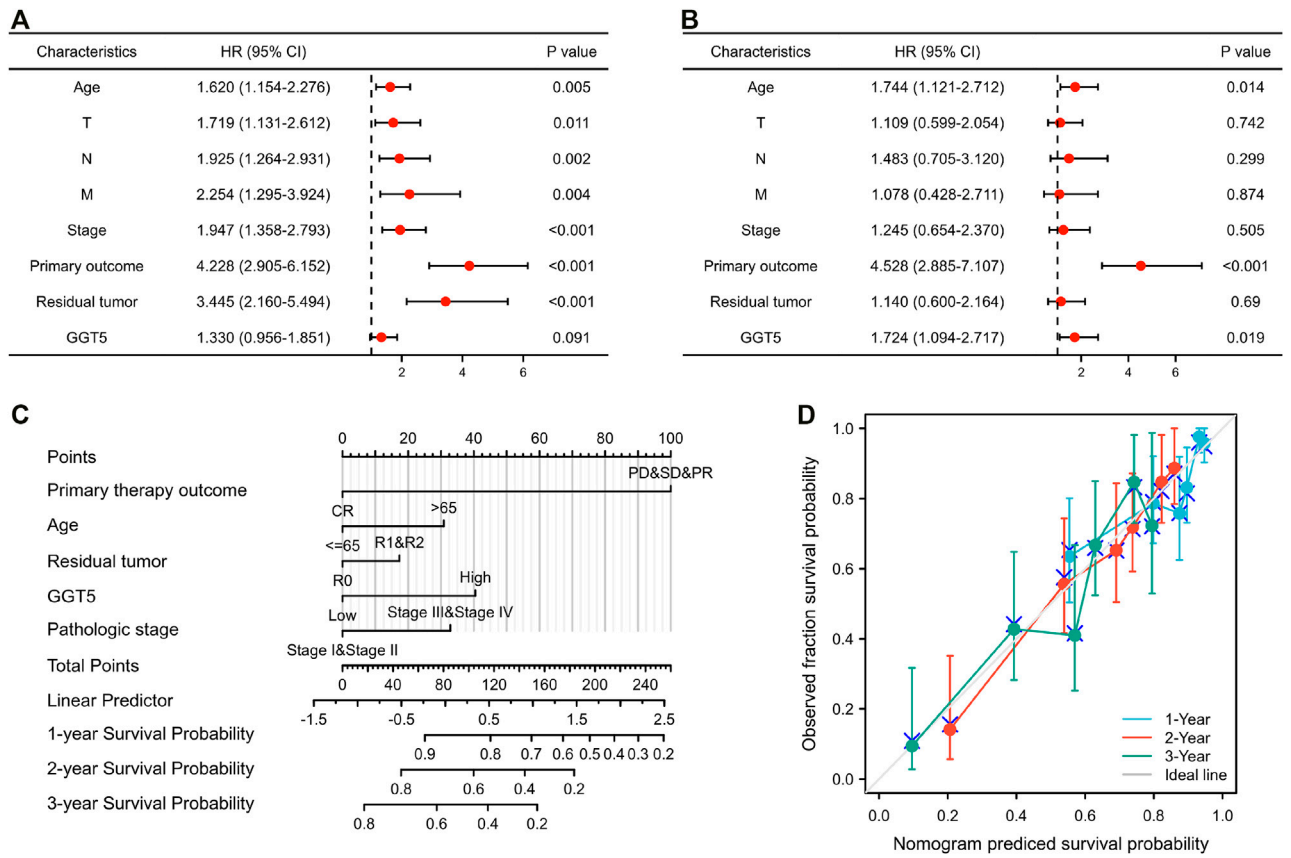


FIGURE 4 | The prognostic factor analysis and nomogram model construction. The results of univariate **(A)** and multivariate Cox regression **(B)** for the overall survival of GC patients are displayed in the form of forest plots. **(C)** A nomogram model was constructed for predicting the 1-year, 2-years, and 3-years overall survival probability of GC patients. **(D)** Calibration curves for the nomogram model predicting the probability of the 1-year, 2-years, and 3-years overall survival of GC patients.

relationship between *GGT5* expression and the degree of immune cell infiltration. The results indicated that high levels of *GGT5* expression were significantly correlated with the high-level infiltration of the majority of immune cells, including T cells, pDC, NK cells, NK CD56dim cells, neutrophils, aDC, B cells, CD_8^+ T cells, cytotoxic cells, DC, eosinophils, iDC, macrophages, mast cells, central memory T-cell (Tcm), effective memory T-cell (Tem), follicular helper T-cell (TFH), T gamma delta (Tgd), Th1 cells, and Treg cells ($p < 0.05$) (**Figure 5A**). Then, the correlation between *GGT5* expression and the immune cell enrichment was analyzed with the Spearman correlation test and determined by the ssGSEA algorithm. *GGT5* expression was positively correlated with the infiltration of NK cells ($r = 0.720$, $p < 0.001$) and macrophages ($r = 0.590$, $p < 0.001$). In contrast, a relatively low level of Th17 infiltration was observed in the high *GGT5* expression group compared with the low *GGT5* expression group. ($r = -0.220$, $p < 0.001$), as shown in **Figures 5B–E**. Additionally, similar to the TCGA cohort, upregulated *GGT5* was positively associated with the infiltration of both NK cells and macrophages in the two GEO validation datasets. However, *GGT5* did not show a significant negative correlation with Th17 in the GSE54129 validation cohort, which may be related to sample size and selection bias, as shown in **Supplementary Figures S1D, S2D.**

Correlation of *GGT5* With Immune-Related Genes and Immune Checkpoint Genes

To further understand the relationship between *GGT5* expression and immune cell infiltration in the microenvironment of gastric cancer, we performed a correlation analysis of *GGT5* and immune-related genes, including MHC genes, immune activation genes, immunosuppressive genes, chemokine receptors, and chemokines. The results indicated that *GGT5* had a positive relationship with MHC genes (**Figure 6A**) and immune activation genes (**Figure 6B**), especially *HLA-DOA*, *CXCL12*, *ENTPD1*, and *STING1*. Remarkably, the majority of immunosuppressive (**Figure 6C**) and chemokine receptors (**Figure 6D**) were positively correlated with *GGT5* expression. In addition, more than half of the chemokines (**Figure 6E**) were positively co-expressed with *GGT5* based on the TCGA cohort, and similar positive correlations of *GGT5* with chemokines were also observed in the validation cohorts of GSE54129 and GSE29272 (**Supplementary Figures S3A–E** and **Supplementary Figures S4A–E**). Therefore, it could be

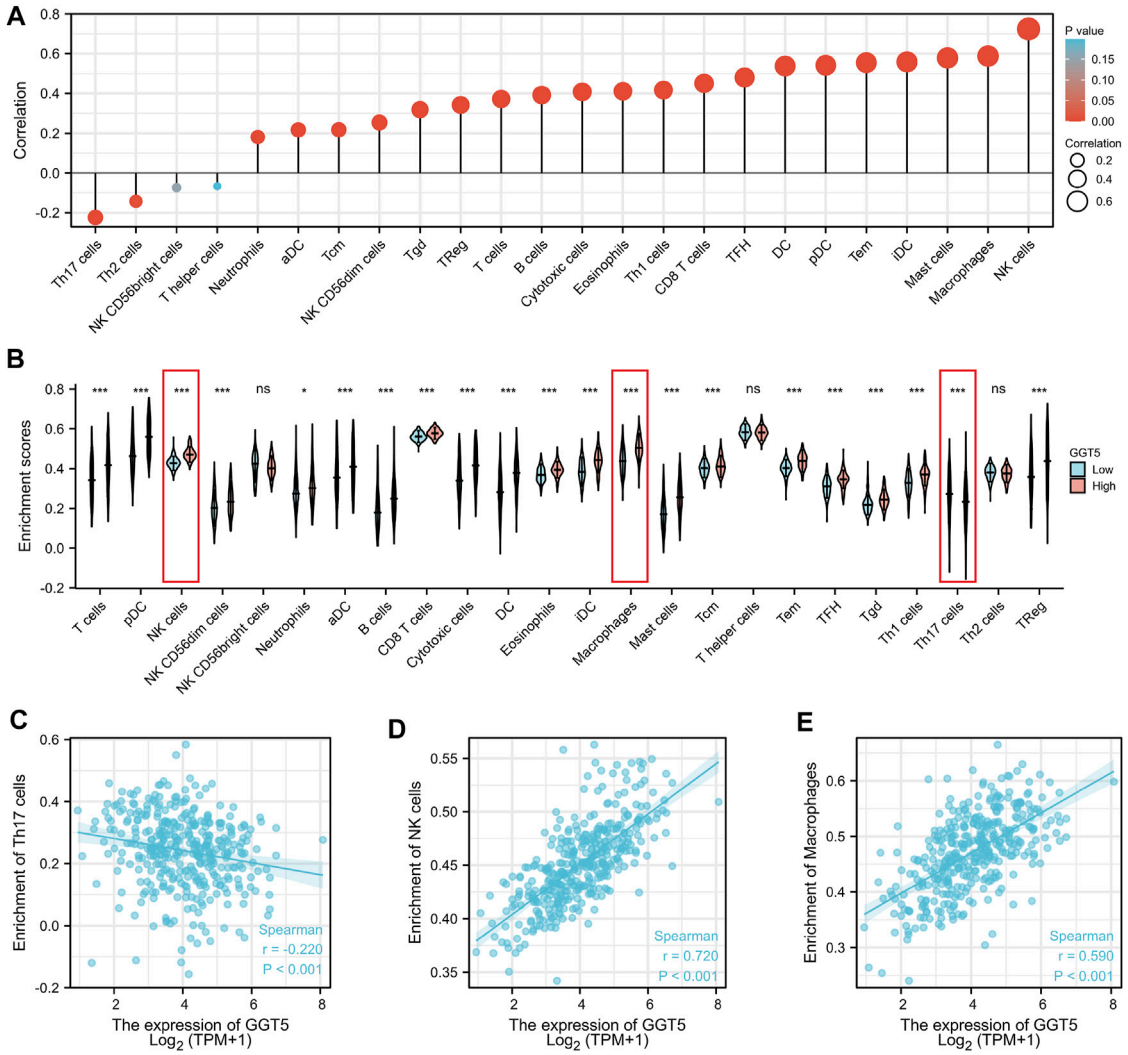


FIGURE 5 | The expression level of *GGT5* was closely associated with the degree of immune cell infiltration in the tumor immune environment. **(A)** Correlation analysis between the expression level of *GGT5* and the degree of immune cell infiltration. **(B)** Comparison of the different immune cells infiltration levels under high and low *GGT5* expression conditions. **(C–E)** The scatter plots show the correlation between *GGT5* expression and the infiltration degrees of Th17 cells, macrophages, and NK cells. *, $p < 0.05$; ***, $p < 0.001$; ns, not significant, $p > 0.05$.

hypothesized that these immune-related genes, especially immunosuppressive and chemokines and related receptors may be involved in the regulation of immune cell infiltration patterns.

Since immunotherapy has been clinically utilized for the treatment of various tumors, we next compared the expression levels of immune checkpoint-related genes in different *GGT5* expression groups. The results showed that in addition to *CD274* (p = 1.25e-01), seven other genes, *CTLA4* (p = 6.53e-04), *HAVCR2* (p = 2.08e-11), *LAG3* (p = 2.37e-03), *PDCD1* (p = 3.88e-06), *PDCD1LG2* (p = 8.24e-15), *TIGIT* (p = 1.13e-07), and *SIGLEC15* (p = 4.09e-02), were obviously upregulated in the *GGT5*-high expression group (**Figures 6F,G**). A similar positive correlation was also seen in the validation dataset of GSE29272 (**Supplementary Figures S4F,G**). However, the expression of *GGT5* did not show a significant positive correlation with these immune checkpoint-related genes in GSE54129, which may be related to the small sample size and selection bias (**Supplementary Figures S3F,G**). Based on the TIDE algorithm to test the clinical response to immune checkpoint blockade, we demonstrated a significantly lower predicted response rate of gastric cancer patients in the *GGT5*-high expression group (**Figure 6H**). Furthermore, co-expression analysis revealed that the expression of *GGT5* was positively associated with these immune checkpoint-related genes in the TCGA cohort (**Figure 6I**), especially for *PDCD1LG2* and *PDCD1*, which were validated in the cohorts of GSE54129 and GSE29272, respectively (**Supplementary Figures S3H, S4H**). All of the above results indicated that *GGT5* may serve as a potential immunotherapy target.

Functional Analysis of *GGT5* in Gastric Cancer

To gain further insight into the potential functions of *GGT5* in gastric cancer, GO categories and KEGG pathway enrichment

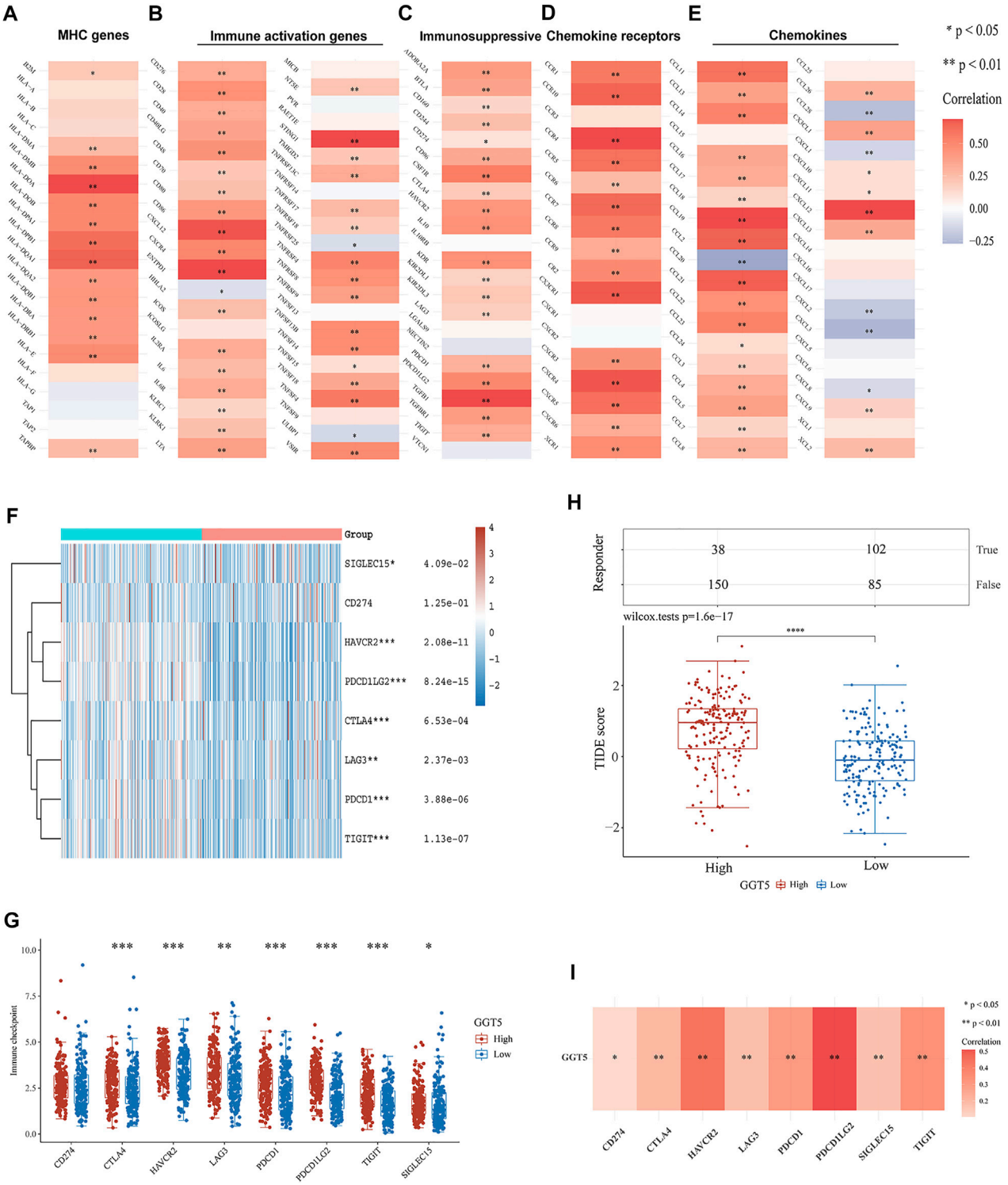


FIGURE 6 | Correlation of *GGT5* with immune-related genes and immune checkpoint genes in gastric cancer. Correlation of *GGT5* and MHC genes **(A)**, immune activation genes **(B)**, immunosuppressive genes **(C)**, chemokine receptors **(D)**, and chemokines **(E)**. **(F,G)** The expression patterns and levels of immune checkpoint genes in different *GGT5* expression groups. **(H)** The predicted response to immune checkpoint blockade therapy in different *GGT5* expression groups. **(I)** Correlation of *GGT5* with immune checkpoint-related genes. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

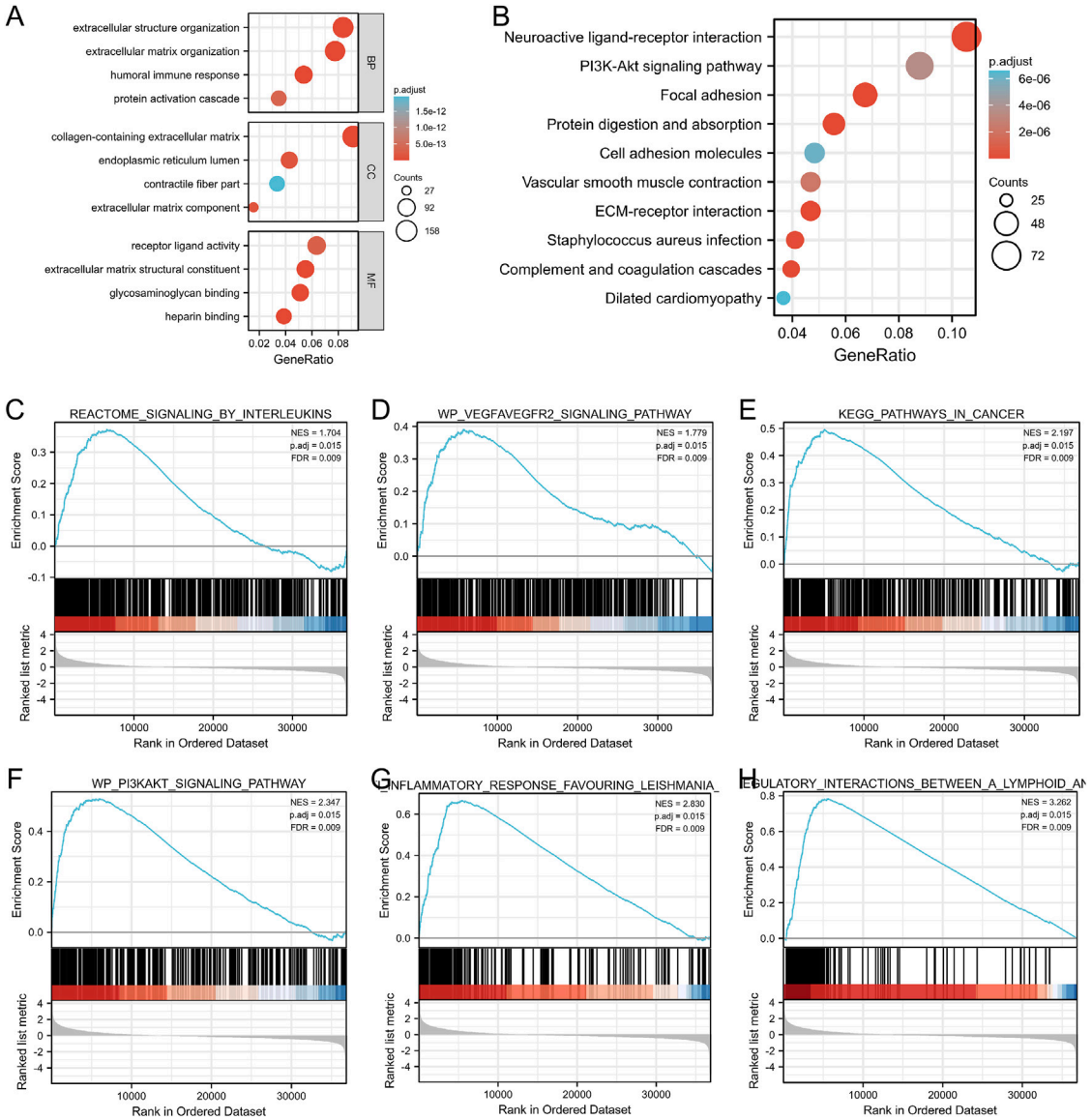


FIGURE 7 | GO categories and pathway enrichment analysis of *GGT5* in gastric cancer. **(A)** GO annotations of DEGs. **(B)** KEGG pathway enrichment analysis determined the top 10 significantly enriched pathways. **(C–H)** GSEA was applied to identify *GGT5*-relevant signaling pathways and biological processes including signaling by interleukins, the VEGFA-VEGF2 signaling pathway, pathways in cancers, the PI3K/Akt signaling pathway, anti-inflammatory response favoring Leishmania, and immunoregulatory interactions between lymphoid and non-lymphoid cells. NES: normalized enrichment score; *p*. adj: adjusted *p value*; FDR: false discovery rate.

analyses were carried out based on the TCGA database. The results showed that for the biological process, these *GGT5*-related DEGs were mainly enriched in extracellular structure/matrix organization, humoral immune response, and protein activation cascade. For the cellular component, DEGs were mainly involved in the collagen-containing extracellular matrix, endoplasmic reticulum lumen, and contractile fiber part. For the molecular function, the DEGs were mainly enriched in the processes of receptor-ligand activity, extracellular matrix structural constituent, and glycosaminoglycan binding (**Figure 7A**, **Supplementary Table S6**). In addition, the neuroactive ligand-receptor interaction, PI3K-Akt, focal adhesion, protein digestion, and protein absorption signaling pathways were also closely correlated with the regulation of *GGT5*-related DEGs (**Figure 7B**, **Supplementary Table S7**). Similar functional annotations, such as extracellular matrix structure constituents, collagen-containing extracellular matrix, glycosaminoglycan binding, and focal adhesion pathways, were also enriched in the two GEO validation cohorts (**Supplementary Figures S1E,F**, **Supplementary Figures S2E,F**, and **Supplementary Tables S8, S9**).

GSEA Identified *GGT5*-Relevant Signaling Pathways

GSEA was performed to ascertain the related signaling pathways of *GGT5* in gastric cancer. The enrichment results indicated that pathways related to tumor proliferation and differentiation as well as the immune and inflammatory response were enriched in the *GGT5* high expression group, including signaling by interleukins, the VEGFA-VEGF2 signaling pathway, pathways in cancers, the PI3K/Akt signaling pathway, the anti-inflammatory response favoring Leishmania, and immunoregulatory interactions between lymphoid and non-lymphoid cells (**Figures 7C–H**, **Supplementary Table S10**). Taken together, *GGT5* is an immune-related gene and may play a critical role in inflammatory responses, angiogenesis, and the tumor immune response to promote the development and progression of gastric cancer.

DISCUSSION

To the best of our knowledge, the correlation between *GGT5* and immune infiltration levels in the tumor microenvironment of gastric cancer has not been previously reported. Herein, the present study confirmed that *GGT5* was upregulated in GC tissues and associated with a poor prognosis of patients with GC. ROC analysis implied that *GGT5* could be applied as a prognostic biomarker with moderate predictive value (AUCs between 0.7–0.8 are considered to be moderate) to distinguish GC tissues from normal tissues. Moreover, highly expressed *GGT5* was proven to be closely associated with a high histological grade, pathological stage, and poor prognosis of gastric cancer patients.

Recent studies have found a correlation between elevated serum levels of *GGT* and various digestive tumors, including pancreatic head carcinoma (Lyu et al., 2021), colorectal carcinoma (Gong et al., 2021), and gallbladder carcinoma (Su et al., 2021). A linear association was also detected between increased levels of *GGT* and cancer-related deaths (Albhaisi and Qayyum, 2021). Furthermore, as a traditional biochemical indicator, *GGT* has been widely used to evaluate the severity of digestive system-related diseases. Thus, the intrinsic connections and underlying mechanisms between *GGT* gene members and digestive tumors deserve further in-depth study.

Among the *GGT* gene family members, at least two members, *GGT1* and *GGT5*, have gained more attention due to their remarkable catalytic activity. Both *GGT1* and *GGT5* are type II membrane glycoproteins, which are mainly involved in the metabolism of tripeptide glutathione and leukotriene C4 (LTC4) (Heisterkamp et al., 2008; Hanigan et al., 2015). However, several differences in the organ/tissue distribution of the two enzymes have been observed. Previous studies have shown that *GGT1* is usually located on the apical surface of many epithelial cells, such as the surface of renal proximal tubules, prostate glands, and salivary glands, while *GGT5* is not localized to a specific region of the cell surface. Specifically, in the liver expressing both GGT1 and *GGT5*, *GGT1* is mainly distributed on the bile canaliculi surface, while *GGT5* is significantly expressed in Kupffer cells. Additionally, *GGT5* was found to play a critical role in converting LTC4 to LTD4 in the spleen, liver, and uterus (Hanigan et al., 2015). Its specific high expression in macrophages and its involvement in the metabolism of inflammatory mediators suggest that *GGT5* plays a crucial role in the regulation of the immune system.

According to the results of pan-cancer analysis, we found great heterogeneity in the expression levels of *GGT5* among various tumor types, which indicated that *GGT5* might play different roles in different tumors. Previous studies have concluded that *GGT5* overexpression is correlated with a poor prognosis in patients with lung cancer, gastric cancer, and colon cancer. A recent study showed that for lung adenocarcinoma (LUAD), *GGT5* was specifically highly expressed in cancer-associated fibroblasts instead of the tumor cells, which was proven to contribute to accelerating tumor cell proliferation and drug resistance in LUAD. Meanwhile, significantly poor OS and PFS were observed in LUAD patients with high *GGT5* expression (Wei et al., 2020). Additionally, a study on B-cell malignancy suggested that *GGT5* was overexpressed in follicular dendritic cells, which suggested a mechanism for B-cell confinement mediated by P2RY8 and the ligand S-geranylgeranyl-L-glutathione (Lu et al., 2019). Our present study has confirmed that *GGT5* plays a critical role in the progression of gastric cancer and could serve as an independent prognostic factor for gastric cancer patients.

Since the potential clinical value of *GGT5* in predicting the outcomes of gastric cancer patients was demonstrated, we established a nomogram including *GGT5* and other clinical characteristics (age, primary therapy outcome, residual tumor, and pathologic stage) according to the results of the univariate and multivariate Cox regressions. Such a scoring approach attempts to provide a more accurate and personalized prognostic assessment for gastric cancer patients by incorporating proven risk factors. Both the calibration curves and the C-index showed good agreement between the predicted and actual observed values for the 1-year, 2-years, and 3-years overall survival. Thus, our constructed nomogram model can be treated as a practical tool for individualized survival assessment of gastric cancer patients.

Because overexpression of *GGT5* was highly associated with a poor prognosis in gastric cancer, we further investigated the correlation between immune cell infiltration and *GGT5* expression to understand the potential mechanisms. It is well known that cells that infiltrate the tumor microenvironment, including immune cells, cancer cells, stromal cells, and extracellular matrix, have an impact on immune evasion, tolerance, and tumor progression. Our findings suggested that the infiltration levels of NK cells and macrophages were increased in the *GGT5*-high expression group, which showed significant correlations. Conversely, a negative correlation was observed between the Th17 infiltration level and *GGT5* expression. It has been confirmed that the function and phenotype of infiltrated NK cells in the tumor microenvironment are impaired, even leading to the dysfunction or exhaustion of NK cells (Zhang et al., 2020; Riggan et al., 2021). A recent study on

non-small cell lung cancer (NSCLC) showed that increased cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) in the tumor microenvironment could suppress the function of dendritic cells (DCs), resulting in immunosuppressive effects (Russick et al., 2020). The major proportions of macrophages infiltrating the tumor microenvironment are tumor-associated macrophages (TAMs) (Noy and Pollard, 2014). TAMs have emerged as a critical factor in promoting tumor progression by generating a complex mixture of inflammatory cytokines, chemokines, and growth factors (Noy and Pollard, 2014). In contrast, studies have demonstrated that the infiltration and function of Th17 cells in the tumor microenvironment are associated with tumor regression and survival improvement in patients diagnosed with epithelial carcinoma (Wilke et al., 2011). Animal experiments have also confirmed that Th17 cells promote the upregulation of CD_4^+ T lymphocytes in the tumor microenvironment, thus inhibiting tumor growth and prolonging survival in a mouse model of pancreatic carcinoma (Gnerlich et al., 2010). Therefore, our results suggest that the poor prognosis in the *GGT5*-high expression group might be attributed to an imbalance in the immune function homeostasis and an impaired anti-tumor immune response.

To gain further insight into the underlying mechanisms by which *GGT5* was correlated with immune cell infiltration, we performed a correlation analysis between *GGT5* and immune-related genes. The results revealed a positive correlation of *GGT5* with these immune-related genes, especially for immunosuppressive and chemokine receptor genes. Therefore, it could be inferred that the poor prognosis in the *GGT5*-high expression group is closely related to the high expression levels of immunosuppressive-related genes and chemokine receptors. The immune checkpoint analysis indicated that gastric cancer patients with upregulated *GGT5* expression showed a higher TIDE score and expression of *CD274*, *CTLA4*, *HAVCR2*, *LAG3*, *PDCD1*, *PDCD1LG2*, and *TIGIT* than patients in the *GGT5*-low expression group. Studies have shown that patients with a higher TIDE score suggest T cell dysfunction in the tumor microenvironment, which is associated not only with poor immune checkpoint blockade treatment but also with poor survival under anti-PD-1 and anti-CTLA4 therapy (Jiang et al., 2018). Therefore, these findings demonstrated that targeting *GGT5* may be a potential strategy for immune checkpoint blockade treatment.

To explore the biological function of *GGT5* in gastric cancer, *GGT5*-related DEGs were screened and subjected to GO functional annotation, GSEA, and KEGG pathway enrichment analyses. The GO functional annotations showed that extracellular matrix, receptor activity, and immune response were significantly enriched based on the TCGA database, and the extracellular matrix was especially enriched in both the TCGA and GEO cohorts. Interestingly, as a major component in the extracellular matrix, glycosaminoglycans are involved in all stages of cancer progression. For instance, cancer cells can secrete glycosaminoglycans such as heparinase and hyaluronidase to penetrate the basement membrane and extracellular matrix to invade surrounding tissues (Yip et al., 2006). Recent studies have shown that glycosaminoglycans can also interact with chemokines and drain biologics with chemokine neutralization functions, leaving free chemokines available to combine with chemokine receptors and promote cancer progression (Ortiz Zacarias et al., 2021). GSEA indicated that signaling by interleukins, the VEGFA-VEGF2 signaling pathway, pathways in cancers, the PI3K/Akt signaling pathway, the anti-inflammatory response favoring Leishmania, and immunoregulatory interactions between lymphoid and nonlymphoid cells were closely correlated with the *GGT5*-high phenotype. Based on the enrichment analyses, we may infer that *GGT5* plays important role in promoting the carcinogenesis and development of gastric cancer via a series of biological processes, such as immune response, angiogenesis, and inflammatory response. These results are consistent with the co-expression analysis of *GGT5* and immune-related genes.

Several studies have demonstrated that IL-6-, IL-8-, IL-10-, and IL-33 mediated pathways are involved in the invasion and metastasis of gastric cancer (Chung and Lim, 2017; Yang et al., 2018; Chen et al., 2019; Ham et al., 2019). For example, as a proinflammatory factor, IL33 was shown to promote the malignant progression of gastric cancer by activating the downstream ST2/MAPK/ERK1/2 signaling cascade (Yu et al., 2015; Zhou et al., 2020; Huang N. et al., 2021), which turned out to be a valuable prognostic biomarker for gastric cancer patients (Sun et al., 2011; Hu et al., 2017). In addition, IL33 could also inhibit platinum-induced apoptosis and promote cell invasion via the ST2/MAPK/JNK pathway, conferring resistance to gastric cancer chemotherapy (Ye et al., 2015). The PI3K/Akt/mTOR signaling pathway is commonly accepted as a vital pathway involved in various human cancers, and mediating epithelial-mesenchymal transformation and chemoresistance is considered to be the major factor promoting gastric cancer progression (Fattahi et al., 2020). Previous studies also revealed that VEGFR-2 mediated the major angiogenic functions of VEGF-A, including stimulating the proliferation and migration of endothelial cells, increasing vascular permeability, and promoting angiogenesis (Park et al., 2015). The circulating VEGE-A levels have been reported to be closely correlated with the treatment response, pathological characteristics, and prognosis of gastric cancer patients. Furthermore, antiangiogenic therapy targeting VEGF-A/VEGFR2, such as bevacizumab and ramucirumab, was recognized as an effective treatment strategy for advanced gastric cancer (Hironaka, 2019).

Our research confirmed the differential expression of *GGT5* between gastric cancer and normal tissues and increased our understanding of *GGT5* and the progression of gastric cancer from the perspective of immune cell infiltration. However, several limitations exist in the present study. First, all samples involved in the current study were based on RNA sequencing data from online databases. In addition to the data heterogeneity and platform differences, lacking or inconsistent clinical information may affect the accuracy of the outcomes. Therefore, prospective studies with large sample sizes are required to confirm our findings. Second, the association between *GGT5* expression and immune cell infiltration as well as the direct molecular mechanism of *GGT5* involvement in the progression of gastric cancer need further validation. Finally, the sample sizes of our study were still

inadequate, particularly the results of some subgroup analyses, which may be affected by random chance. Additional studies with larger sample sizes are needed to prove these findings.

CONCLUSION

Taken together, we confirmed that *GGT5* was highly expressed in gastric cancer tissues compared to normal samples, and it could be identified as a specific biomarker for distinguishing gastric cancer tissues from normal gastric mucosa. Upregulated *GGT5* was proven to be closely associated with poor overall survival and progression-free intervals in gastric cancer patients and could be applied as a clinically independent prognostic factor. A nomogram model was further constructed and computationally validated for individualized overall survival assessment. The immune cell infiltration analysis showed that *GGT5* expression was positively correlated with NK cells and macrophages but negatively correlated with the infiltration of Th17 cells. Additionally, we revealed that *GGT5* was co-expressed with immune-related genes and immune checkpoint genes and that *GGT5* may be a potential target for immune checkpoint blockade treatment. Finally, functional annotation and pathway enrichment analysis supported that *GGT5* was mainly involved in the biological processes of the immune response, angiogenesis, and inflammatory response. Our research provides a novel perspective for further understanding the mechanisms by which *GGT5* is correlated with immune cell infiltration in the tumor microenvironment of gastric cancer.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

YW and YF contributed equally to this study. All the listed authors directly, substantially, and intellectually contributed to the preparation of this manuscript and approved the publication.

FUNDING

This study was supported by the National Natural Foundation of China (no. 81973795 to LY) and the Shanghai Municipal Natural Science Foundation (19ZR1452200 to FZ).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.810292/full#supplementary-material

Supplementary Figure S1 | Independent validation of *GGT5*-related functional annotation, pathway enrichment, and its relationship with immune cell infiltration in GSE54129. **(A)** ROC curve showing the predictive value of *GGT5* for identifying gastric cancer tissues based on GSE54129. **(B,C)** Volcano plot **(B)** and heatmap clustering **(C)** of *GGT5*-related DEGs. **(D)** Correlation analysis between *GGT5* expression and immune cells infiltration. **(E)** GO annotations of *GGT5*-related DEGs. **(F)** KEGG pathway enrichment analysis displayed the significantly enriched pathways.

Supplementary Figure S2 | Independent validation of *GGT5*-related functional annotation, pathway enrichment, and its relationship with immune cell infiltration in GSE29272. **(A)** ROC curve showing the predictive value of *GGT5* for identifying gastric cancer tissues based on GSE29272. **(B,C)** Volcano plot **(B)** and heatmap clustering **(C)** of *GGT5*-related DEGs. **(D)** Correlation analysis between *GGT5* expression and immune cells infiltration. **(E)** GO annotations of *GGT5*-related DEGs. **(F)** KEGG pathway enrichment analysis displayed the significantly enriched pathways.

Supplementary Figure S3 | Validation of *GGT5*'s correlation with immune-related genes and immune checkpoint genes based on GSE54129. Correlation of *GGT5* and MHC genes **(A),** immune activation genes **(B),** immunosuppressive genes **(C),** chemokine receptors **(D),** and chemokines **(E)**. **(F,G)** The expression patterns and levels of immune checkpoint genes in different *GGT5* expression groups. **(H)** Correlation of *GGT5* expression level with immune checkpoint-related genes. *, $p < 0.05$; **, $p < 0.01$.

Supplementary Figure S4 | Validation of *GGT5*'s correlation with immune-related genes and immune checkpoint genes based on GSE29272. Correlation of *GGT5* and MHC genes **(A),** immune activation genes **(B),** immunosuppressive genes **(C),** chemokine receptors **(D),** and chemokines **(E)**. **(F,G)** The expression patterns and levels of immune checkpoint genes in different *GGT5* expression groups. **(H)** Correlation of *GGT5* expression level with immune checkpoint-related genes. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

Supplementary Table S1 | List of *GGT5* mRNA expression levels in gastric cancer and normal gastric tissues in GEO and TCGA databases.

Supplementary Table S2 | List of all differentially expressed genes.

Supplementary Table S3 | List of 330 significantly differentially expressed genes with the absolute log2 fold change (FC) > 2 and adjusted *p*-value < 0.05 as cutoff criteria identified from the TCGA database.

Supplementary Table S4 | List of 507 significantly differentially expressed genes detected from GSE54129.

Supplementary Table S5 | List of 38 significantly differentially expressed genes detected from GSE29272.

Supplementary Table S6 | List of GO enrichment analyses of DEGs based on the TCGA database.

Supplementary Table S7 | List of KEGG enrichment analysis results based on the TCGA database.

Supplementary Table S8 | List of GO annotations and KEGG pathway enrichment analysis based on GSE54129.

Supplementary Table S9 | List of GO annotations and KEGG pathway enrichment analysis based on GSE29272.

Supplementary Table S10 | List of gene set enrichment analysis results.

REFERENCES

Abrahao-Machado, L. F., and Scapulatempo-Neto, C. (2016). HER2 Testing in Gastric Cancer: An Update. *World J. Gastroenterol.* 22, 4619–4625. doi:10.3748/wjg.v22.i19.4619

Albhaisi, S., and Qayyum, R. (2021). The Association between Serum Liver Enzymes and Cancer Mortality. *Clin. Exp. Med.* 22, 75. doi:10.1007/s10238-021-00733-9

Balachandran, V. P., Gonen, M., Smith, J. J., and Dematteo, R. P. (2015). Nomograms in Oncology: More Than Meets the Eye. *Lancet Oncol.* 16, e173–e180. doi:10.1016/s1470-2045(14)71116-7

Bang, Y.-J., Van Cutsem, E., Feyereislova, A., Chung, H. C., Shen, L., Sawaki, A., et al. (2010). Trastuzumab in Combination with Chemotherapy versus Chemotherapy Alone for Treatment of HER2-Positive Advanced Gastric or Gastro-Oesophageal junction Cancer (ToGA): a Phase 3, Open-Label, Randomised Controlled Trial. *Lancet* 376, 687–697. doi:10.1016/s0140-6736(10)61121-x

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2012). NCBI GEO: Archive for Functional Genomics Data Sets-Update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193

Bergquist, J. R., Leiting, J. L., Habermann, E. B., Cleary, S. P., Kendrick, M. L., Smoot, R. L., et al. (2019). Early-onset Gastric Cancer Is a Distinct Disease with Worrisome Trends and Oncogenic Features. *Surgery* 166, 547–555. doi:10.1016/j.surg.2019.04.036

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A. C., et al. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity* 39, 782–795. doi:10.1016/j.immuni.2013.10.003

Cappellesso, R., Fassan, M., Hanspeter, E., Bornschein, J., S.G. d'Amore, E., Cuorvo, L. V., et al. (2015). HER2 Status in Gastroesophageal Cancer: a Tissue Microarray Study of 1040 Cases. *Hum. Pathol.* 46, 665–672. doi:10.1016/j.humpath.2015.02.007

Chen, L., Shi, Y., Zhu, X., Guo, W., Zhang, M., Che, Y., et al. (2019). IL-10 Secreted by Cancer-associated Macrophages Regulates Proliferation and Invasion in Gastric Cancer Cells via c-Met/STAT3 Signaling. *Oncol. Rep.* 42, 595–604. doi:10.3892/or.2019.7206

Chen, G., Chen, D., Feng, Y., Wu, W., Gao, J., Chang, C., et al. (2022). Identification of Key Signaling Pathways and Genes in Eosinophilic Asthma and Neutrophilic Asthma by Weighted Gene Co-expression Network Analysis. *Front. Mol. Biosci.* 9, 805570. doi:10.3389/fmolb.2022.805570

Cheng, Y., Xing, S. G., Jia, W. D., Huang, M., and Bian, N. N. (2017). Low PLCE1 Levels Are Correlated with Poor Prognosis in Hepatocellular Carcinoma. *Oncol. Targets Ther.* 10, 47–54. doi:10.2147/OTT.S126401

Chung, H. W., and Lim, J. B. (2017). High-mobility Group Box-1 Contributes Tumor Angiogenesis under Interleukin-8 Mediation during Gastric Cancer Progression. *Cancer Sci.* 108, 1594–1601. doi:10.1111/cas.13288

Fattahi, S., Amjadi-Moheb, F., Tabaripour, R., Ashrafi, G. H., and Akhavan-Niaki, H. (2020). PI3K/AKT/mTOR Signaling in Gastric Cancer: Epigenetics and beyond. *Life Sci.* 262, 118513. doi:10.1016/j.lfs.2020.118513

Gnerlich, J. L., Mitchem, J. B., Weir, J. S., Sankpal, N. V., Kashiwagi, H., Belt, B. A., et al. (2010). Induction of Th17 Cells in the Tumor Microenvironment Improves Survival in a Murine Model of Pancreatic Cancer. *J. Immunol.* 185, 4063–4071. doi:10.4049/jimmunol.0902609

Gong, Z., Zhang, X., Geng, Q., Li, W., Huang, M., Chen, Z., et al. (2021). AKP and GGT Level Can Provide an Early Prediction of First-Line Treatment Efficacy in Colorectal Cancer Patients with Hepatic Metastases. *Biomarkers Med.* 15, 697–713. doi:10.2217/bmm-2020-0667

Ham, I.-H., Oh, H. J., Jin, H., Bae, C. A., Jeon, S.-M., Choi, K. S., et al. (2019). Targeting Interleukin-6 as a Strategy to Overcome Stroma-Induced Resistance to Chemotherapy in Gastric Cancer. *Mol. Cancer* 18, 68. doi:10.1186/s12943-019-0972-8

Hanigan, M. H., Gillies, E. M., Wickham, S., Wakeham, N., and Wirsig-Wiechmann, C. R. (2015). Immunolabeling of Gamma-Glutamyl Transferase 5 in normal Human Tissues Reveals that Expression and Localization Differ from Gamma-Glutamyl Transferase 1. *Histochem. Cel Biol.* 143, 505–515. doi:10.1007/s00418-014-1295-x

Heisterkamp, N., Groffen, J., Warburton, D., and Sneddon, T. P. (2008). The Human Gamma-Glutamyltransferase Gene Family. *Hum. Genet.* 123, 321–332. doi:10.1007/s00439-008-0487-7

Hironaka, S. (2019). Anti-angiogenic Therapies for Gastric Cancer. *Asia Pac. J. Clin. Oncol.* 15, 208–217. doi:10.1111/ajco.13174

Hu, W., Wu, C., Li, X., Zheng, Z., Xie, Q., Deng, X., et al. (2017). Serum IL-33 Level Is a Predictor of Progression-free Survival after Chemotherapy. *Oncotarget* 8, 35116–35123. doi:10.18632/oncotarget.16627

Huang, N., Cui, X., Li, W., Zhang, C., Liu, L., and Li, J. (2021a). IL-33/ST2 Promotes the Malignant Progression of Gastric Cancer via the MAPK Pathway. *Mol. Med. Rep.* 23, 361. doi:10.3892/mmr.2021.12000

Huang, X., You, S., Ding, G., Liu, X., Wang, J., Gao, Y., et al. (2021b). Sites of Distant Metastases and Cancer-specific Survival in Intraductal Papillary Mucinous Neoplasm with Associated Invasive Carcinoma: A Study of 1,178 Patients. *Front. Oncol.* 11, 681961. doi:10.3389/fonc.2021.681961

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T Cell Dysfunction and Exclusion Predict Cancer Immunotherapy Response. *Nat. Med.* 24, 1550–1558. doi:10.1038/s41591-018-0136-1

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Lu, E., Wolfreys, F. D., Muppidi, J. R., Xu, Y., and Cyster, J. G. (2019). S-Geranylgeranyl-L-glutathione Is a Ligand for Human B Cell-Confinement Receptor P2RY8. *Nature* 567, 244–248. doi:10.1038/s41586-019-1003-z

Lyu, S.-C., Wang, J., Huang, M., Wang, H.-X., Zhou, L., He, Q., et al. (2021). CA19-9 Level to Serum γ-Glutamyltransferase as a Potential Prognostic Biomarker in Patients with Pancreatic Head Carcinoma. *Cancer Manag. Res.* 13, 4887–4898. doi:10.2147/cmar.s313517

Noy, R., and Pollard, J. W. (2014). Tumor-Associated Macrophages: From Mechanisms to Therapy. *Immunity* 41, 49–61. doi:10.1016/j.immuni.2014.06.010

Ortiz Zacarías, N. V., Bemelmans, M. P., Handel, T. M., De Visser, K. E., and Heitman, L. H. (2021). Anticancer Opportunities at Every Stage of Chemokine Function. *Trends Pharmacol. Sci.* 42, 912–928. doi:10.1016/j.tips.2021.08.001

Park, D. J., Thomas, N. J., Yoon, C., and Yoon, S. S. (2015). Vascular Endothelial Growth Factor a Inhibition in Gastric Cancer. *Gastric Cancer* 18, 33–42. doi:10.1007/s10120-014-0397-4

Ren, J., Feng, J., Song, W., Wang, C., Ge, Y., and Fu, T. (2020). Development and Validation of a Metabolic Gene Signature for Predicting Overall Survival in Patients with colon Cancer. *Clin. Exp. Med.* 20, 535–544. doi:10.1007/s10238-020-00652-1

Riggan, L., Shah, S., and O'sullivan, T. E. (2021). Arrested Development: Suppression of NK Cell Function in the Tumor Microenvironment. *Clin. Transl. Immunol.* 10, e1238. doi:10.1002/cti2.1238

Russick, J., Joubert, P.-E., Gillard-Bocquet, M., Torset, C., Meylan, M., Petitprez, F., et al. (2020). Natural Killer Cells in the Human Lung Tumor Microenvironment Display Immune Inhibitory Functions. *J. Immunother. Cancer* 8, e001054. doi:10.1136/jitc-2020-001054

Su, S., Liu, L., Sun, C., Nie, Y., Guo, H., Hu, Y., et al. (2021). Preoperative Serum Gamma-Glutamyltransferase as a Prognostic Biomarker in Patients Undergoing Radical Cystectomy for Bladder Cancer. *Front. Oncol.* 11, 648904. doi:10.3389/fonc.2021.648904

Sun, P., Ben, Q., Tu, S., Dong, W., Qi, X., and Wu, Y. (2011). Serum Interleukin-33 Levels in Patients with Gastric Cancer. *Dig. Dis. Sci.* 56, 3596–3601. doi:10.1007/s10620-011-1760-5

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Van Cutsem, E., Bang, Y.-J., Feng-Yi, F., Xu, J. M., Lee, K.-W., Jiao, S.-C., et al. (2015). HER2 Screening Data from ToGA: Targeting HER2 in Gastric and Gastroesophageal junction Cancer. *Gastric Cancer* 18, 476–484. doi:10.1007/s10120-014-0402-y

Wang, G., Hu, N., Yang, H. H., Wang, L., Su, H., Wang, C., et al. (2013). Comparison of Global Gene Expression of Gastric Cardia and Noncardia Cancers from a High-Risk Population in china. *PLoS One* 8, e63826. doi:10.1371/journal.pone.0063826

Wei, J.-R., Dong, J., and Li, L. (2020). Cancer-associated Fibroblasts-Derived Gamma-Glutamyltransferase 5 Promotes Tumor Growth and Drug Resistance in Lung Adenocarcinoma. *Aging* 12, 13220–13233. doi:10.18632/aging.103429

Wen, F., Huang, J., Lu, X., Huang, W., Wang, Y., Bai, Y., et al. (2020). Identification and Prognostic Value of Metabolism-Related Genes in Gastric Cancer. *Aging* 12, 17647–17661. doi:10.18632/aging.103838

Wickham, S., West, M. B., Cook, P. F., and Hanigan, M. H. (2011). Gamma-glutamyl Compounds: Substrate Specificity of Gamma-Glutamyl Transpeptidase Enzymes. *Anal. Biochem.* 414, 208–214. doi:10.1016/j.ab.2011.03.026

Wilke, C. M., Kryczek, I., Wei, S., Zhao, E., Wu, K., Wang, G., et al. (2011). Th17 Cells in Cancer: Help or Hindrance? *Carcin* 32, 643–649. doi:10.1093/carcin/bgr019

Wu, J., Wang, X., Wang, N., Ma, L., Xie, X., Zhang, H., et al. (2021). Identification of Novel Antioxidant Gene Signature to Predict the Prognosis of Patients with Gastric Cancer. *World J. Surg. Oncol.* 19, 219. doi:10.1186/s12957-021-02328-w

Yang, T., Zhang, J., Zhou, J., Zhu, M., Wang, L., and Yan, L. (2018). Resveratrol Inhibits Interleukin-6 Induced Invasion of Human Gastric Cancer Cells. *Biomed. Pharmacother.* 99, 766–773. doi:10.1016/j.biopha.2018.01.153

Ye, X.-L., Zhao, Y.-R., Weng, G.-B., Chen, Y.-C., Wei, X.-N., Shao, J.-P., et al. (2015). IL-33-induced JNK Pathway Activation Confers Gastric Cancer Chemotherapy Resistance. *Oncol. Rep.* 33, 2746–2752. doi:10.3892/or.2015.3898

Ye, Z., Zheng, M., Zeng, Y., Wei, S., Huang, H., Wang, Y., et al. (2021). A 13-Gene Metabolic Prognostic Signature Is Associated with Clinical and Immune Features in Stomach Adenocarcinoma. *Front. Oncol.* 11, 612952. doi:10.3389/fonc.2021.612952

Yip, G. W., Smollich, M., and Götte, M. (2006). Therapeutic Value of Glycosaminoglycans in Cancer. *Mol. Cancer Ther.* 5, 2139–2148. doi:10.1158/1535-7163.mct-06-0082

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Yu, X.-X., Hu, Z., Shen, X., Dong, L.-Y., Zhou, W.-Z., and Hu, W.-H. (2015). IL-33 Promotes Gastric Cancer Cell Invasion and Migration via ST2-ERK1/2 Pathway. *Dig. Dis. Sci.* 60, 1265–1272. doi:10.1007/s10620-014-3463-1

Zhang, J., Zhao, T., Han, F., Hu, Y., and Li, Y. (2019). Photothermal and Gene Therapy Combined with Immunotherapy to Gastric Cancer by the Gold Nanoshell-Based System. *J. Nanobiotechnol.* 17, 80. doi:10.1186/s12951-019-0515-x

Zhang, C., Hu, Y., and Shi, C. (2020). Targeting Natural Killer Cells for Tumor Immunotherapy. *Front. Immunol.* 11, 60. doi:10.3389/fimmu.2020.00060

Zhou, Q., Wu, X., Wang, X., Yu, Z., Pan, T., Li, Z., et al. (2020). The Reciprocal Interaction between Tumor Cells and Activated Fibroblasts Mediated by TNF-α/IL-33/ST2L Signaling Promotes Gastric Cancer Metastasis. *Oncogene* 39, 1414–1428. doi:10.1038/s41388-019-1078-x

Zhu, C., Xiao, H., Jiang, X., Tong, R., and Guan, J. (2021a). Prognostic Biomarker DDOST and its Correlation with Immune Infiltrates in Hepatocellular Carcinoma. *Front. Genet.* 12, 819520. doi:10.3389/fgene.2021.819520

Zhu, H., Hu, X., Gu, L., Jian, Z., Li, L., Hu, S., et al. (2021b). TUBA1C Is a Prognostic Marker in Low-Grade Glioma and Correlates with Immune Cell Infiltration in the Tumor Microenvironment. *Front. Genet.* 12, 759953. doi:10.3389/fgene.2021.759953

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership